

Methods to Identify and Characterize Developmental Neurotoxicity for Human Health Risk Assessment. I: Behavioral Effects

Deborah A. Cory-Slechta,¹ Kevin M. Crofton,² Jeffery A. Foran,³ Joseph F. Ross,⁴ Larry P. Sheets,⁵ Bernard Weiss,¹ and Beth Mileson⁶

¹Department of Environmental Medicine, University of Rochester Medical School, Rochester, New York, USA; ²U.S. Environmental Protection Agency, National Health and Environmental Effects Research Laboratory, Research Triangle Park, North Carolina, USA; ³Citizens for a Better Environment, Milwaukee, Wisconsin, USA; ⁴The Procter & Gamble Company, Ross, Ohio, USA; ⁵Toxicology Department, Bayer Corporation, Stilwell, Kansas, USA; ⁶ARCADIS Geraghty & Miller, Millersville, Maryland, USA

Alterations in nervous system function after exposure to a developmental neurotoxicant may be identified and characterized using neurobehavioral methods. A number of methods can evaluate alterations in sensory, motor, and cognitive functions in laboratory animals exposed to toxicants during nervous system development. Fundamental issues underlying proper use and interpretation of these methods include *a*) consideration of the scientific goal in experimental design, *b*) selection of an appropriate animal model, *c*) expertise of the investigator, *d*) adequate statistical analysis, and *e*) proper data interpretation. Strengths and weaknesses of the assessment methods include sensitivity, selectivity, practicality, and variability. Research could improve current behavioral methods by providing a better understanding of the relationship between alterations in motor function and changes in the underlying structure of these systems. Research is also needed to develop simple and sensitive assays for use in screening assessments of sensory and cognitive function. Assessment methods are being developed to examine other nervous system functions, including social behavior, autonomic processes, and biologic rhythms. Social behaviors are modified by many classes of developmental neurotoxicants and hormonally active compounds that may act either through neuroendocrine mechanisms or by directly influencing brain morphology or neurochemistry. Autonomic and thermoregulatory functions have been the province of physiologists and neurobiologists rather than toxicologists, but this may change as developmental neurotoxicology progresses and toxicologists apply techniques developed by other disciplines to examine changes in function after toxicant exposure. *Key words:* behavioral testing, cognitive function, developmental neurotoxicity, motor activity, sensory function. — *Environ Health Perspect* 109 (suppl 1):79–91 (2001).

<http://ehpnet1.niehs.nih.gov/docs/2001/suppl-1/79-91/cory-slechta/abstract.html>

The normal structure and function of the nervous system may be altered as a result of exposure to some xenobiotics before or after birth. Alterations in nervous system function may be identified in laboratory animals using neurobehavioral methods. Our understanding of neurobehavioral methods is derived primarily from an extensive history of use in four related disciplines: experimental psychology, ethology, biopsychology, and behavioral pharmacology. Because of the vast array of technologies and experimental models provided by these histories, the neurobehavioral panel of the working group convened by the International Life Sciences Institute Risk Science Institute decided to limit and focus our discussion to those most relevant and promising for developmental assessment. We divided the available assessment methods into six categories: sensory function, motor function, cognitive function, social behaviors, autonomic/thermoregulatory processes, and biologic rhythms. Developmental neurotoxicity (DNT) data available for the first three categories far exceed the data available for the latter three; therefore, discussion of the first three categories was more extensive and focused on basic principles that form the

basis for proper use and interpretation. The consensus of the neurobehavioral panel was that the behavioral test methods used in DNT testing are, for the most part, employed correctly in hundreds of different laboratories around the world. However, there are numerous examples of the misuse of these methods and misinterpretation of results derived from these methods. Therefore, a major focus of the discussion that follows outlines the principles for proper use and interpretation of these methods. The latter three categories were discussed less extensively and primarily from the standpoint of potential usefulness for DNT. Although neurophysiologic techniques also have an extensive history of use in both neuroscience and clinical neurology (*1*), these have not been used on a wide scale for DNT studies and were outside the scope of the current discussion.

Common Issues

Many issues are common to most, if not all, methods of behavioral testing for DNT. The importance of most of these issues has been recognized, with attempts to ensure that methodology addresses them appropriately. Discussions in this section provide additional

guidance on consideration of these issues in the use of behavioral test methods in DNT testing.

Behavioral tests vary along many dimensions of desirable properties, including the amount of available validation data, speed of testing, breadth and/or specificity of test results, availability of equipment and personnel to conduct the test, and extrapolation of results among species. Thus, the most desirable properties depend upon the experimental context (i.e., what is the question being asked? What other end points are available to address the issue?). The latter question is extremely important and often overlooked because most tests are part of a battery. If little is known about a test substance and the investigator is screening for an effect, the most desirable properties of a test battery may be a wide breadth of function(s) tested, relatively short testing period, low cost, and availability of personnel. In contrast, specificity and sensitivity of effect are more desirable test attributes if the chemical is known to produce weakness, for example, and a second tier test is used to characterize the effect and to determine a no-observed adverse effect level, i.e., distinguish between diminished ability to exert high forces versus ataxia. In the latter case, cost of the test and wide availability of personnel may be less important.

Animal Model

The choice of animal models in developmental neurotoxicology studies is influenced by a

Address correspondence to B. Mileson, ARCADIS Geraghty & Miller, 1131 Benfield Blvd., Millersville, MD 21108 USA. Telephone: (410) 987-0032. Fax: (410) 987-4392. E-mail: bmileson@arcadis-us.com

This project was supported by a cooperative agreement between the International Life Sciences Institute Risk Science Institute and the U.S. Environmental Protection Agency Office of Pesticide Programs, by the National Institute for Environmental Health Sciences, and by the International Life Sciences Institute Research Foundation.

The opinions expressed herein are those of the individual authors and do not necessarily reflect the views of their respective organizations or the sponsoring institutions. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Received 5 September 2000; accepted 13 December 2000.

number of factors. Each of these factors may have more or less weight depending upon the ultimate scientific goal. Choice of the animal model should be determined primarily by the hypothesis being tested. The species and strain of animal should be appropriate for the target system being tested or modeled. For example, albino strains of laboratory animals are not appropriate when modeling effects of chemicals on human vision, as albino rodents typically have extremely poor vision.

In studies conducted for risk assessment purposes, extrapolation of animal data to humans may lead to a different choice. The chicken is the animal model of choice when testing for organophosphate-induced delayed neuropathy, primarily because of the predictive power of the resulting data (2). Economics may also factor into a decision of which animal model to use. If screening unknown agents is the major driving need, lower-cost assays may allow one to evaluate many more compounds. Screening chemicals in a common rodent species such as the mouse or rat is more cost effective than use of nonhuman primates. If economics are a deciding factor, one needs to be convinced that the resulting data will still be meaningful. Age can be a crucial factor in determining the correct animal model. For example, testing a hypothesis concerning the role of exposure to pesticides as a risk factor for Parkinson's disease would necessitate the need to employ lifetime studies (or other models appropriate for testing mechanisms of aging). If enough information is available on the mechanism of action of a class of chemicals, transgenic or congenic animal models may be useful.

Although not a prerequisite for choosing an animal model, adequate background information on the normal anatomy, physiology, and behavior of the test species can be useful in interpretation of toxicant-induced changes. Historical control data are also useful in this regard. Last, the generality of the test method, or concurrent validity, should be established by determining whether the effect correlates with other indices of toxicity. For example, do effects observed in a behavioral test of olfactory function correlate with the underlying pathologic damage to the olfactory epithelium in the nasal cavity and/or olfactory bulbs in the central nervous system?

Age relevance. Design of behavioral experiments and hypotheses should consider the age relevance of the animal model. The onset and maturation of most behaviors are necessarily linked to the age of the animal. Although not always critically evaluated, age relevance of the test procedure can be very important in determining the validity of a test method for use in DNT studies. Use of procedures validated in adult models may not be appropriate for young animals.

Resource Demands

Regulatory testing requires that many behavioral techniques be able to test large numbers of animals (e.g., 10–20 litters per treatment group, with four to five treatment groups) in a relatively rapid fashion. Procedures that do not lend themselves to rapid testing can lead to an inability to test enough subjects at an appropriate age. It is important to minimize personnel costs directly related to time and effort spent testing animals; however, cost should not be an excuse for lack of testing. Instead, excessive costs associated with a specific test method should be a clarion call for development of more cost-effective tests.

Many companies now make available complete, turn-key, computer-based systems to assess specific behavioral functions (e.g., water mazes for measurement of learning/memory, shuttle-box avoidance for memory, locomotor activity chambers, startle testing equipment). In many cases, the software and hardware components of these systems tend to be simplistic and inflexible. Several issues should be considered before such equipment is acquired. For example, a misperception about the real costs of carrying out tests of cognitive functions often leads investigators to purchase less expensive equipment in an attempt to economize. The less automated the device, the more time required of the investigator/staff to conduct the behavioral tests, e.g., putting animals into start boxes, measuring times, errors or other dependent variables, and housing animals during intertrial intervals. Thus, although single-use equipment may cost less to acquire originally, the economic resources to use the equipment may actually be far greater than starting with automated, flexible equipment. In addition, the use of hardware/software dedicated to measurement of a specific behavioral function does not guarantee that it measures that behavioral function selectively, and means that additional hardware/software will have to be acquired for every other behavioral function the investigator ultimately wishes to measure. In contrast, operant chambers, which may have a greater up-front cost, may at the same time provide greater automation and flexibility of use.

Expertise and Training

There often appears to be a misperception that implementation and interpretation of behavioral tests are easily accomplished. To the contrary, as in any area of science, expertise and training in behavioral sciences are critical for both. The absence of such training often results in a lack of understanding of the variables that may influence a behavior being measured and failure to adequately control for the impact of these variables on behavior. Additionally, absence of appropriate training and expertise in behavioral methods often

leads to inappropriate interpretation of outcome measures in cognitive tests. This has become apparent in the appearance in the literature of studies examining cognitive function in genetically engineered mice. In this case, state-of-the-art molecular biologic approaches may be used in conjunction with misuse and misinterpretation of simple tests of learning or memory. Such occurrences reflect the lack of expertise and training of the investigators in this dimension of their experiments and the absence of expertise on the advisory boards of these journals.

Many people involved in assessing behavioral and/or neurologic function in toxicity studies have some training in experimental psychology or psychopharmacology as well as statistics. This training is needed to ensure proper study design, data collection, and data analysis. Behavioral tests of sensory end points are sensitive to a wide variety of environmental variables (e.g., ambient noise, handling history, time of day) that may not appear important to untrained investigators.

Many of the more sophisticated and sensitive cognitive measures may require taking the subjects through a series of training programs, in which expertise is again critical for precisely molding behavior to that required for the final stage of assessment. Likewise, the conduct of special sensory tests (e.g., reflex audiometry) requires adequate training in sensory psychophysics, statistics, and the principles of behavioral testing. However, there is no systematic training or certification program comparable to those available for personnel involved in recording and interpreting other important toxicity end points using anatomic or clinical pathologic techniques (3). Thus, there is considerable variability in the expertise of individuals generating and interpreting behavioral and/or neurologic data in academic, industrial, or contract laboratories. The inherent problems associated with the lack of professional standards are compounded by the wide availability of off-the-shelf equipment that may appear to obviate the need for expertise. It must be stressed that behavioral testing devices are tools that depend on the expertise of the user. Proper selection and use of the tools require expertise. Moreover, the design, conduct, and interpretation of studies including behavioral and/or neurologic end points require personnel with relevant training and experience. For example, personnel who examine neurologic end points including reflex and reaction, spontaneous movement abnormalities, and open-field changes in gait and posture should have adequate training in the conduct of these tests and use of appropriate terminology (4). In addition, efforts should be made to ensure consistency among observers, and interobserver reliability should be reported.

Statistics

A number of statistical issues should be considered in developmental neurotoxicology studies. The first issue concerns how best to control for litter effects. Littermates may be assigned to more than one task, individual animals may be repeatedly tested on a task, or different littermates may be tested on different tasks. The statistical implications differ for these testing strategies. Because most developmental studies use the litter as the unit of measure, repeated sampling from that litter (using more than one animal from each litter) poses unique statistical problems. Simply put, use of more than one animal per litter inflates the number of subjects per group and can increase type I error (i.e., false positives). There are a number of good reviews of this subject that should be consulted prior to designing studies (5–8).

Repeated measures are also a common factor in DNT testing. Current U.S. Environmental Protection Agency guidelines (9) recommend motor activity testing in the same animal at least 3 times during the preweaning period. When analysis of variance (ANOVA)-based statistics are used, repeated-measures data must be treated as such in the model. Alternative measures such as regression techniques also can be used (10). Finally, the assignment of litter members to one or more tasks must be carefully considered. Litter representatives can be selected and used for all tests, or separate animals can be assigned to each test. In the former case, behavioral history is a serious confounder. For example, the same animals should not be used in learning and memory tests at both younger and older ages.

Variability is an inherent property of all test methods and population samples. The greater the variability among animals, the larger the number required to detect an effect. A variety of techniques can be used to determine, a priori, the power of a test method based on historical control data, and thus predict the number of animals needed to detect an effect of a specific size (11,12). Use of these statistical procedures is strongly advised.

In the case of negative findings, especially in regulatory testing, documentation of historical control data and positive control data is important. These data are necessary to adequately document the power of the test and thus assure a low incidence of type II error. Historical databases are encouraged to be made available or published whenever possible [for example, see Crofton et al. (13)]. Laboratories employing test methods for the first time are encouraged to compare their data variability with published reports and/or historical control data [for example, see Wise et al. (14)]. Test methods that result in extremely high variability compared to that in

other laboratories should be inspected to determine the source of the variability or not be used. Positive control data, although not as important, are needed to establish changes in the end point related to dose response whenever possible. Dose–response data (and not single-dose studies) should be stressed, as these data are needed to adequately document the ability to detect different magnitudes of effect. Dose–response data are also needed to determine the linearity or nonlinearity of the effect.

System-Specific Issues

Although some issues are common to most if not all DNT behavioral testing methods, many methods and their desirable properties are unique to specific functions. Further, some functions or behavioral parameters have not been the focus of DNT testing; thus, methods are only poorly developed or remain undeveloped. This section presents desirable properties of methods for DNT testing in common behavioral functions (e.g., sensory, motor, cognitive). It also presents a discussion of social behavior, and the issues to be considered as DNT testing methods are developed for this area.

Sensory Function

Sensation plays a crucial role in the ability of the organism to interact with its environment. Loss of some aspect of sensory function is one of the most common occupational injuries (15,16), and approximately 44% of all neurotoxic chemicals are reported to affect sensory function adversely (15,17,18). Although the exact magnitude of the problem may be debatable, there are two major reasons that sensory systems should be evaluated in the safety assessment of potential neurotoxicants. First, most sensory organs are not protected by the blood–brain barrier and may experience greater exposure than other parts of the nervous system. This is especially true for the olfactory system. Second, proper interpretation of the results from other types of behavioral function (e.g., cognitive testing) requires consideration of alterations in sensory system processing as a confounder.

Assessing sensory function in animals has been a significant research tradition in comparative psychology, and many of the methods developed in that field have been applied in neurotoxicology. These methods vary from relatively simple and subjective sensory reflex tests, such as elicitation of the pinna reflex and pupil constriction, to more complicated operant-discrimination paradigms and evoked potential procedures. The major sensory systems of concern in toxicology include visual, auditory, olfactory, nociception (pain and other noxious stimuli), somatosensory, and vestibular. The methods used to study these different systems, by

necessity, will be somewhat different. However, the characteristics of the properties of the test methods that should be under experimental control will be very similar [for review, see Maurissen (19)].

Animal model. Species and strain of the test animal are important in tests of sensory systems. Selection of the animal model should be determined primarily by the hypothesis or specific aims. First and foremost should be the use of a species/strain appropriate for the sensory system being tested or modeled. For example, use of some strains of mice (e.g., C57Bl/6J) may be a poor choice for auditory studies because of the early-onset presbycusis found in this strain (20). However, use of the rat as an animal model of ototoxicity has been highly successful in modeling the adverse effects of xenobiotics in humans (21–23). Stebbins and colleagues, in a series of now-classic studies in this area, used monkeys and guinea pigs to model the dose response and time course of the ototoxicity of aminoglycoside antibiotics (24–26). An excellent example of the simultaneous assessment of multiple sensory systems comes from the work of Pryor et al. (27). These authors used a conditioned avoidance procedure to assess the effects of a wide variety of chemicals on both auditory and visual system functions in rats.

An important issue in developmental toxicity testing is that the animal model selected must also be age relevant. For example, stimulus parameters (e.g., nociceptive stimuli) using a conditioned lick-suppression paradigm for adult animals may be inappropriate in much younger animals. Procedures with long acquisition times (e.g., operant) may not be able to target rapidly maturing sensory systems. The work of Merigan and colleagues (28), which characterized the adverse effects of acrylamide on visual function in monkeys, used operant methods that would be unsuitable for testing the ontogeny of visual function in rats because of extensive training demands. Lastly, stimuli may not be perceived, or may be perceived as less intense, in animals with immature sensory systems. For example, sonalerts and click stimuli are commonly used auditory stimuli in adult testing. However, these generate low-frequency stimuli that would be inappropriate in preweaning rats or mice, as sensation of low frequencies is the last to develop in most altricial rodents (29,30). An appropriate understanding of the normal ontogeny of the sensory system being assessed is necessary to ensure a good match between the age of the animal and the test procedure.

Any discussion of animal models should also include the issue of sensitivity. There are numerous sensory techniques that are simple and inexpensive, but *a*) may lack precision in estimating psychophysical thresholds, *b*) may

be subject to experimenter bias, *c*) are difficult to automate, or *d*) are generally result in large interlaboratory variability (19). Reflexive movements (e.g., pinna reflex, corneal reflex) are examples of simple and inexpensive end points to test. These tests are fast, easy, and inexpensive. However, they are thought to be relatively insensitive to toxicant-induced alterations in sensory function (31,32). Conversely, operant and conditioned discrimination procedures represent some of the more sensitive and specific methods used for assessing sensory system dysfunctions. Such tests have been used to characterize the effects of a wide variety of agents that disrupt various sensory functions. Examples include auditory deficits produced by aminoglycoside antibiotics (25), visual deficits produced by methyl mercury (33), ammonia-induced disruption of olfaction (34), and somatosensory dysfunction produced by acrylamide (35). Performance of these tests generally requires extensive training of the subject. These tests are time consuming and generally are limited to small numbers of subjects. Collection of data required to characterize the audiogram for the rat using a conditioned lick-suppression paradigm required over 150 days of testing (36). Reflex modification audiometry has been used as a fairly rapid and sensitive test of auditory function [for example, see Crofton et al. (37) and Young and Fetcher (38)]. This procedure also works in humans [for review, see Ison (39)]. However, this procedure requires equipment not available as a turn-key system from commercial suppliers. The trade-off between economy and sensitivity may be difficult to balance for some sensory function testing.

Stimulus parameters. The proper generation and use of stimuli are crucial for testing the effects of xenobiotics on sensory system function. A number of stimulus properties are shared by all sensory systems, including intensity, frequency, duration, and location in space. When using sensory methods in toxicity testing, it is extremely important that normative data be developed for a method that demonstrates the ability of the test method to detect and characterize the effects of varying the magnitude of such variables. For example, in tests of visual function, response magnitude should be directly related to the intensity of the visual stimulus (40). Auditory testing thresholds should vary across the frequency domain (41). Stimulus frequency should also be appropriate to the test species. For example, use of a 0.5-kHz tone as the conditioning stimulus in rats (42) is not optimal, as rats do not hear this frequency very well (36). An 8-kHz stimulus would be more appropriate for use in rats because the threshold for this frequency is approximately 45 dB lower than a 0.5-kHz stimulus.

Stimulus amplitude may be too high to detect small changes in thresholds. For example, routine DNT testing of auditory function using startle habituation uses a high-decibel stimulus (e.g., 110 dB sound pressure level). A false-negative finding would result should animals demonstrate only a small increase in threshold for that stimulus. Recent work with exposure to polychlorinated biphenyls during development has found only small changes (i.e., 15–25 dB) in low-frequency thresholds (37). These changes would not be detected in tests using high-decibel stimuli.

Statistics. Variability can arise from different sources in tests of sensory function. It is inherent in any test species (43) and is a property of the test method itself or in its application. Any potentially confounding variable not controlled will lead to increased variance. The rapid development in young animals should preclude averaging data across ages, as the response could change drastically in a very short time. In DNT testing, averaging startle response data in animals tested over a few days of age in adults will not necessarily add to the group variance. However, in young animals, averaging response data over several days could easily increase variance due to increases in body weight and increases in sensitivity of the developing auditory system.

All tests of sensory function should be able to generate data that have adequate statistical power to detect biologically relevant changes in behavior. For example, a 15-dB change in auditory thresholds is generally regarded as adverse (44). Therefore, any method used should be capable of detecting statistical differences between group means that differ by 15 dB or more. Historical control data and positive control data should be available to document this property adequately. Positive control data should establish changes in the end point related to the dose response whenever possible.

Analyses and interpretation. Interpretation of the results of sensory assessment tests should follow the classic rules of behavioral science. First, sensory testing never yields a direct measurement of sensation. Instead, one infers a change in sensory function based on the observed change in the motor response being evaluated. For example, increased latency to paw lick in the hot-plate test is a behavioral measure that is interpreted, after ruling out other causes, as an increase in a nociceptive threshold (45). Motor impairment due to muscle fiber degeneration may also lead to increases in latency that are not caused by sensory mechanisms. Administration of a high dose of a sedative, with subsequent decrease in the amplitude of the acoustic startle response, does not mean that the sedative induces an auditory dysfunction. Instead, decreases in motor capability are more likely

the culprit for this effect [see Maurissen (19) for a review of the use and misuse of psychophysical methods].

Interpretation of any treatment-related change should be done in concert with an understanding of the ability of the test method to detect changes in the response. Data from positive control studies can yield much information on the sensitivity of a test method, as the method is employed in the laboratory generating the data. Data demonstrating experimental responsiveness to changes in stimulus parameters are also necessary. A method that fails to detect differences in stimulus strength in control animals is being either inappropriately employed or inappropriately interpreted. Historical control data are also invaluable in this regard. Excessive variability in control group means over time may be indicative of a lack of proper experimental control of the behavior being assessed. Alternatively, individual differences due to a bimodal effect in a population are possible. Reliability of the test method will, of course, depend on a finding of low variability in control values over time, as well as replication in the effects of positive control agents.

Research needs. Included in the research needed to advance our ability to determine sensory system toxicity is a better understanding of the relationship between the simple sensory system tests currently used in DNT studies (i.e., simplex reflex tests, startle habituation) and any underlying changes in the structure and/or function of the sensory system. A commonly held belief is that simple reflex tests are not as sensitive as tests of sensory thresholds or sensory signal processing (32). Very few studies have tested this hypothesis (31). In addition, current DNT test batteries do not routinely assess the age of onset of sensory function. Without this type of testing, delays in development of sensory systems will go undetected routinely. The question of whether these delays lead to long-term adverse effects in the organism is also currently unanswered. There is also a need to identify and characterize the impact of various potential confounds on the ontogeny and function of sensory systems. What effect do changes in maternal nutritional status, which may be due to chemical exposure, have on measurement of function in offspring? Where do screening methods fit into a tiered testing scheme? This is an especially vexing problem for sensory testing, as the sensitivity of the more rapid test currently used in tier 1 testing is unknown. Current sensory function testing usually evaluates treatment-related effects solely by measuring the behavioral response to amplitude changes in the stimuli. Many types of stimulus processing are not normally assessed (e.g., frequency, duration, and spatial

information). These stimulus properties reflect different degrees of sensitivity to the effects of xenobiotics. Boyes and colleagues (46), using evoked potential, demonstrated deficits in spatial contrast sensitivity with no apparent change in the amplitude of flash-evoked potentials. Last, behavioral toxicologists need to keep abreast of recent advances in neurobiology and genomics that may allow an increased understanding of the physiologic and structural bases for sensory function. This information will be crucial in updating methods in the future.

Motor Function

In toxicity studies the goal of motor function testing is to detect and/or characterize motor dysfunction. Behavioral tests of motor dysfunction in animals are differentiated into two types: those that detect spontaneous movement disorders such as changes in gait, tremors, and myoclonus; and those that detect changes in induced movement such as reflexes, reactions, and movements under operant control. Some end points are recorded subjectively, using categorical (present/absent) or ordinal (e.g., absent, minimal, moderate, severe) scales. For these end points, the data are based upon the judgment of the tester, much as a veterinary neurologist evaluates a patient. Quantitative procedures can be used to measure qualities such as forelimb or hindlimb grip strength. In this case, a transducer detects the response of the subject, and a value is recorded.

A review of all the behavioral tests of motor function is beyond the scope of this section, which will be limited to those categories of tests most commonly used to evaluate motor function in experimental and regulated DNT studies. These categories include observation of locomotion, measurement of locomotor activity, and tests of reflexes and reactions.

Observation. Observation of locomotion is used primarily to detect changes in posture and gait (e.g., ataxia, low carriage) and spontaneous movement abnormalities (e.g., stereotypy, myoclonus, tremors). Observational techniques are fast and inexpensive and enable detection and characterization of a wide variety of functional changes (47). The correct application of observational techniques requires a level of training and experience sometimes not fully appreciated (3), and appropriate use of terminology is necessary to derive maximum usefulness of this approach (4). The sensitivity of qualitative observations is not clear but is generally thought to be less than that of quantitative procedures.

Motor activity measurement. The term motor activity refers to a wide variety of tests that measure different aspects of behavior, for example, exploration, navigation, and

emotionality (48,49). At least three technical aspects must be considered to appreciate the similarities and differences among test systems for measuring locomotor activity: *a*) the size (i.e., small or large), shape (e.g., square, round, or figure eight), and illumination (i.e., dark, dim, bright); *b*) the transducer or detector device (e.g., photocells or video camera); and *c*) the statistical analysis applied to the raw data. The data analyses include macroanalysis (e.g., path length, percent of movement on interior or exterior of environment), microanalysis [e.g., quantification of individual behaviors such as turning, rearing, sniffing; (50)], and characterization of path shape [e.g., predictable vs unpredictable (51)]. The test animal may be placed in the middle of a symmetric device and its spontaneous behavior measured, or the environment may be enhanced asymmetrically with objects that present distinctive visual or tactile stimuli. In the latter case, the orientation of the subject or attention to the stimuli may be evaluated (52).

In DNT studies one common approach is to measure activity using photocells in a small and unenhanced chamber. The behavior evaluated by a particular photocell or combination of photocells may vary by location in the chamber. For example, in some devices a lower row measures horizontal movement (i.e., ambulation), and an upper row measures vertical movements (i.e., rearing). It is essential that the tester understand the relationship between photocell location and behavioral specificity before trying to interpret the results obtained by automated equipment.

Level of activity, reported as number of photocell counts, is the end point most commonly reported, analyzed, and interpreted. The pattern of movements both within session and between sessions may be characterized. Patterns observed within a session may reflect habituation, and patterns observed between sessions may reflect evolution of activity over time, for example, during days 13–21 of postnatal development. Locomotor activity results sometimes have high variability, particularly in developmental studies when animals mature at different rates. The interpretation of motor activity values is generally more controversial than observations of movement disorders (53–55). The algorithm for relating changes in motor activity to neuropathologic end points is not as clear as it is for movement disorders.

Reflexes and postural reactions. Tests of elicited motor function most often include reflex and reaction tests, which can be measured qualitatively or quantitatively. A reflex is an involuntary and relatively stereotyped response to a specific sensory stimulus. The location and amplitude of the response depends on both the location and strength of

the stimulus. For spinal reflexes (e.g., flexor reflex), the sensory stimuli arise from receptors in muscles, joints, and skin, and the neural circuitry responsible for the motor response is contained entirely in the spinal cord. Homologous cranial nerve reflexes are contained within the brain. Although the neuronal circuits that mediate reflexes are simple, the brain frequently coordinates the action of several reflex circuits to generate more complex behaviors that clinical neurologists term reactions (e.g., placing reaction). Reflex and reaction tests are generally quick and easy to perform and require modest training and expertise. Procedures used to quantify some of the reflexes and reactions include forelimb/hindlimb grasp, auditory startle, and extensor thrust (56,57).

Postural reactions are complex responses that maintain the normal, upright position of an animal under conditions of shifting loads. If the weight of an animal is shifted from one side to the other, from front to rear, or from rear to front, the increased load on the supporting limb or limbs requires increased tone in the extensor muscles to keep the limb from collapsing. Part of the alteration in tone is accomplished through spinal reflexes, but for the changes to be smooth and coordinated, the sensory and motor systems of the brain must be involved.

Abnormalities of complex postural reactions (e.g., hopping) do not provide as anatomically precise information about neurologic abnormalities as do reflex tests, which are more limited in scope. However, the intense demands on functional performance required by tests of postural reactions may reveal deficits in neurologic components that are not detected simply by observing gait. The basis for the deficits may then be clarified by further testing of individual reflexes or by electrodiagnostic testing.

Two popular procedures for evaluating motor function—the hindlimb splay and rotorod tests—require special discussion. The hindlimb splay test is conducted by holding a rat horizontally several centimeters above a table surface, then measuring the interpaw distance of the hind limbs after dropping the rat to the table surface. This test is popular because of its sensitivity in detecting acrylamide neurotoxicity (58). However, despite its popularity and its inclusion in the U.S. EPA neurotoxicity testing guidelines (59), little is known about this test. The anatomic basis for the test is unknown, there is no obvious analog used by veterinary or human neurologists, and the neurologic basis for the test can only be hypothesized. The rationale for using the width of the response, as in index of function, is not at all clear. Although an increase in width has been reported for animals

treated with acrylamide, interpretation of a decrease in width is less clear.

In the rotorod test a rat is placed on a slowly rotating rod or dowel, and the end point is the duration that the rat maintains its perch on the rod (60). Although commercially available the rotorod has achieved only modest popularity, primarily because it is time-consuming and cumbersome to conduct, and some animals jump spontaneously from the rod. The rotorod test illustrates the premise that off-the-shelf equipment does not substitute for expertise and training (see section "Expertise and Training").

Analysis and interpretation. In adult neurotoxicity studies the strength of a reflex or reaction is an important index of function. Because the strength of the response is most frequently measured, it is easy to overlook the fact that all reflexes and reactions have a sensory component. For example, failure to flex the leg in response to a toe pinch can reflect loss of sensation or inability to move the leg. Thus, patterns of changes in several reflexes and reactions are generally characterized by neurologic examination to provide interpretable data. In developmental studies the evolution of a reflex or reaction over time is particularly important. Delays in the appearance of a reflex or reaction are important indicators of an adverse developmental effect. The reflexes and reactions most commonly evaluated in developmental studies include flexor reflex, extensor thrust, rooting, placing, surface and air righting, grasp reflex, auditory startle, negative geotaxis, and swimming (61,62).

Research needs. A current dilemma in assessing motor function is that tests of reflexes and reactions are relatively quick and easy to perform, but their sensitivity in the context of DNT testing is uncertain. In contrast, complex procedures exist for characterizing even minor changes in some motor functions [for example, see Stanford and Fowler (63)], but the scope of such procedures is limited, and the procedures too cumbersome to be used routinely. Thus, there is a need to develop and validate technology and procedures that measure motor function objectively and sensitively, yet are sufficiently flexible to be used with large numbers of animals.

There is an additional need to develop technology to measure motor activity. Evaluating the behavior of rats in a novel environment has fascinated and frustrated scientists in a variety of disciplines, including ethology, experimental psychology, psychopharmacology, neuroscience, and neurotoxicology (48,53–55,64,65). Decades ago, observers of behavior recognized that rats in novel environments engaged in a variety of behaviors that suggested this testing environment might be used to evaluate sensory and motor function, emotions, and/or cognition

(66–68). Indeed, contemporary experimentation has shown that the movements of rats in a novel environment reflect the activity of a coordinated navigational system that depends on allothetic (e.g., visual, olfactory, tactile stimuli) and ideothetic (e.g., proprioceptive stimuli from the animal's own movement) stimuli (69–70). The pattern of behavior also reflects the emotional state of the rat (71,72) as well as motor function (73). Recently, observation of rat behavior in a novel environment has emerged as a principal tool for detecting the effects of neurotoxicants on a wide range of neurologic functions (3,74).

Availability of this rich collection of behaviors that rats exhibit in a novel environment has prompted both scientists and commercial equipment manufacturers to develop technology to measure behavior in this test situation (38,73,75,76). The two most common types of systems are based on photocell (75) or video (77) technology. To date, no automated test system has been capable of satisfactorily detecting and quantifying even a small percentage of the range of normal and abnormal behavioral functions that are available and that can be obtained through careful observation. The initial richness of behavior that stimulates interest in open-field activity is also the challenge to be overcome in designing test paradigms. Specifically, a number of diverse normal and abnormal behaviors can occur, and it is valuable to obtain a sense of both the temporal and spatial distribution of the behavior. More complicated technology systems with multiple subsystems (e.g., photocell, video, ultrasound, touch detectors, proximity detectors) or data analysis protocols have been developed (51,76,78–81), but none has been sufficiently useful or practical to become universally popular.

Cognitive Function

The nature of the experimental question should guide the pursuit of the behavioral baselines that will be used. For example, simple tests of learning, providing they are well controlled, may provide information about whether a potential deficit in cognitive function occurs (detection of effect), whereas more sensitive and complex procedures may provide information about the behavioral mechanisms by which such effects occur (characterization of effect). Reliance on complex approaches to answer questions about potential cognitive deficits could be useful, even in screening, to determine whether a potential learning/memory impairment might be attributable to deficits in other areas of nervous system function.

The current U.S. EPA protocol for DNT testing (9) requires assessment of cognitive function at two ages. Certainly such measures are critical components of a DNT assessment

to address concerns over potential long-term consequences of exposures to toxicants during periods of brain development. When assessing cognitive function, numerous issues must be considered before implementing such measures.

Cognitive function is often thought of as encompassing learning, memory, and attention processes. Both learning and memory functions have been extensively studied, defined, and described in the behavioral neuroscience literature. The issue of attention lags behind, as it has not been as systematically studied and defined. The term attention remains a global behavioral construct that may include numerous response classes such as distractibility, impulsivity, sensitivity to delay, activity level, perseveration, sustained attention, and inability to manage delay of reward.

Animal model. The learning task selected should be one appropriate to the species and to the developmental age of the subjects being tested. For example, many different paradigms have been developed to assess aspects of cognitive function, but these have different applicability across species as well as across developmental stages of life. One strategy that may be advantageous to the process of risk assessment is the use of the same behavioral paradigms across species, including humans (82). Tests such as repeated learning or acquisition of response chains and delayed matching paradigms, respectively, can be used across species with appropriate parametric modifications. These are well-validated approaches that have been used extensively and thus allow incorporation of a large database into the assessment. Using these methods eliminates the need for the assumption that dependent variables from neuropsychologic and clinical tests in humans measure the same behavioral processes as the experimental cognitive testing procedures.

Stimulus parameters. Most tests of cognitive function employ measures of accuracy as a primary dependent variable. One important aspect of the test is the level of accuracy maintained under normal conditions. The task should maintain levels of accuracy from which either increases or decreases as a result of exposures can be measured. Tasks that are too easy, as indicated by the ability of subjects to achieve high levels of accuracy, are not sufficiently sensitive for detecting toxicant effects. Carson et al. (83) reported, for example, that lead-exposed sheep exhibited no difference from control in learning to discriminate a vertical from a horizontal line gradient, whereas they exhibited lower levels of accuracy in discriminating a large from a small circle, with the latter discrimination requiring significantly longer to learn in controls. Similarly, Wood et al. (84) demonstrated that toluene disrupted behavior that was at lower accuracy

levels but did not disrupt behavior maintained at high accuracy levels in a fixed consecutive number schedule of reinforcement. Similarly, behavioral paradigms that are too difficult, i.e., those that maintain very low levels of accuracy ultimately, also decrease the probability of detecting a toxicant-induced alteration. Other important aspects of stimulus parameters that must be considered include the saliency of any environmental stimuli used in the cognitive problem to be evaluated and the relevance of the dimension to the species being tested. For example, stimuli that differ in color may fail to be sufficiently discriminable to rodents that possess no color vision (85). Although odors are particularly relevant to mice and rats, the inability of the investigator to precisely control odor onset and offset limits the utility of this measurement and could inadvertently change the nature of the task contingencies (86). Stimulus parameters of other aspects of such tests are also critical (e.g., delay values in memory tests, times between trials), and the literature should be consulted for appropriate values.

Statistics. For most tests of cognitive function, a primary measure of interest will be either accuracy or latency. Ideally, the behavioral tests used to evaluate learning, memory, and attention should provide baseline data for these dependent measures with minimal variability both between subjects and within subjects across time, such that either increases or decreases in these measures can be detected with typical group sizes as reported in the literature. In most

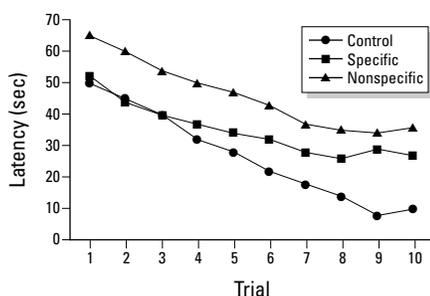


Figure 1. Hypothetical scheme depicting latency in seconds across trials in a water maze in which the task is to locate a submerged platform that permits an escape response. Typically, as acquisition or learning occurs, the latency to find the platform declines across trials (control curve). However, acquisition can be affected by changes in sensory function, motor capabilities, and motivation, all resulting in nonspecific changes in learning. This would typically be manifest as a parallel curve (nonspecific), where differences from control in latency were evident even in trial 1. A treatment-related change in learning that might be indicative of a specific effect is shown as well (specific), in which one would expect the same latency as untreated control in initial trials (indicative of equivalent motor, sensory, and motivational levels), with gradually diverging latencies that do not decline as rapidly as control.

studies of cognitive function, behavior of individual animals is measured repeatedly across sessions (Figures 1,2). When multiple data points are derived from a single subject, they are not considered independent replications; clearly, the behavior of an individual animal is expected to be related to its past performance. Therefore, unless all relevant data are collapsed to a single number, initial statistical analyses require repeated measures approaches that consider this lack of independence. Thus, the common practice of using *t*-tests or one-factor ANOVAs (see example in Figure 3) to analyze multiple data points from single subjects is inappropriate. Repeated measures analyses establish whether there are any main effects of the treatment factor per se as collapsed across the repeated measure. Main effects in the analysis indicate that the behavioral data of the control and treated groups are fit by parallel lines in the simplest case. In other cases the data of control versus treated groups may

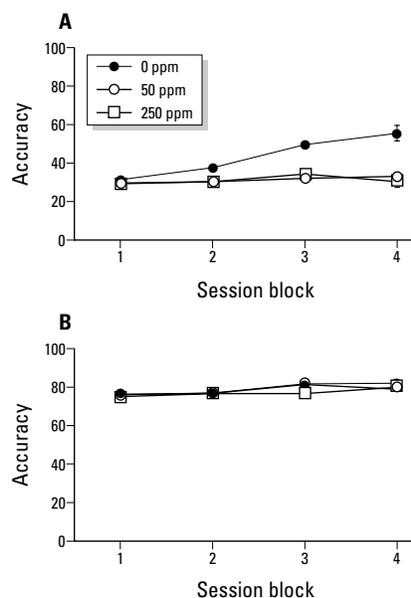


Figure 2. Data from control (0 ppm) and lead-treated (50 ppm, 250 ppm) rats working on a multiple schedule of repeated learning (A) and performance (B). The repeated learning component, which required rats to learn a new sequence of responses during each session, alternated during each session with the performance component, which required only repetition of an already-learned sequence. Accuracy levels during the performance component were substantially higher than in the repeated learning component across sessions, indicative of the development of a strategy for solving the correct sequence for the session. Lead-treated rats showed no such increase and basically remained just above chance levels of accuracy. In contrast, lead-treated rats showed no difficulty in performing an already-learned sequence, thus demonstrating a selective effect of a treatment on learning. Data modified from Cohn et al. (97).

differ only under some circumstances, e.g., only during the final five sessions of testing. In such a case a statistical interaction would be expected from an analysis, which would indicate that there was not only an effect of treatment but that it also occurred only under specific conditions of the repeated factor. With this type of statistical interaction, one predicts intersecting functions of the fits of the control versus treated groups. Only if it can be established that there is either a main effect of treatment or some type of interaction, once the repeated measures have been taken into account across the repeated measure, is it permissible to begin to compare specific data points.

Age relevance. Questions regarding cognitive deficits in response to exposures early in development often focus on long-term outcome, i.e., changes in these behavioral functions as organisms mature. There may be circumstances, however, in which it is desirable to determine in juvenile animals whether cognitive functions have been affected by exposure early in development. Assessing cognitive function in young animals or children requires tests designed to account for the physiologic and physical limitations of the stage of development. While such procedures have been reported in the experimental

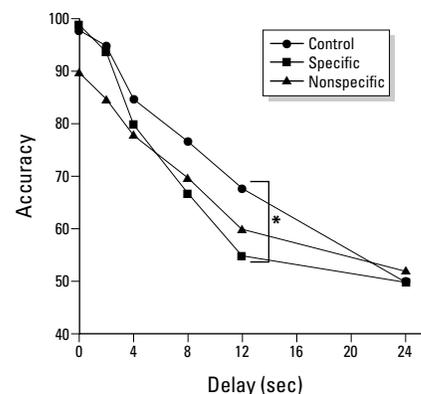


Figure 3. The delayed match-to-sample procedure imposes a delay between a sample stimulus and the subsequent presentation of two stimuli, one of which matches that sample. The selection of the matching stimulus is rewarded. This figure depicts a hypothetical scheme relating changes in accuracy to the length of the delay (in seconds). Typically, accuracy declines as the length of the delay is increased (control). A specific change in memory is indicated by the curve labeled "specific," where accuracy levels of the treated group are equivalent to controls at the 0-sec delay, where no delay is imposed and no remembering required. When accuracy levels are lower even in the 0-sec delay (nonspecific), it indicates that other changes in behavior (e.g., motor, sensory, motivational) are contributing to the deficit. Asterisk (*) and bracket ([]) indicate a significant difference in accuracy between "specific" and "control" at the 12-sec delay point only, based on the inappropriate use of a *t*-test rather than the appropriate repeated measures approach.

animal literature (87–90), they have yet to be fully incorporated into standard test batteries, although the DNT battery requires testing at around 24 days of age and again at about 60–70 days of age. When animal assignments are made in a DNT study for cognitive tests, behavioral histories of the test subjects should be considered. In most cases, animals tested at a young age should not be tested using the same method at the later age. Previous learned behaviors may carry over into the second testing, confounding any ability to assess learning at the later age. Care should be taken to fully understand the ramifications of this. Further, some of the operant touch-screen technologies used in studies with adult humans are being used increasingly in studies with children as young as 4–5 years of age (91–93). This is encouraging, but further development and validation of such tests are urgently needed.

Analysis and interpretation. In many simpler tests of learning and memory, it is possible for changes in sensory function, motor behavior, and/or motivation to indirectly influence the dependent measures that are used and therefore change behavior. Simple paradigms generally fail to include control procedures for assessing these possibilities. Although the water maze is a popular method to measure learning and has proven useful in many contexts (94–96), it also provides numerous examples of the difficulty of interpreting learning impairments. For example, deficits in motor behavior such as strength, endurance, or coordination might result in increased swimming times required to reach an escape platform in a water maze. As decreases in latencies are considered an index of learning in this paradigm, the longer latencies could be misinterpreted as a learning impairment. In a water maze this might result in the type of hypothetical data presented in Figure 1 (nonspecific difference), where the control and nonspecific groups exhibit parallel decreases in latency over the course of trials in such a task. The notable difference in latency, even in the first trial, would suggest that noncognitive influences were produced by treatment and contribute to the differences between the curves. A function more consistent with an interpretation of specific changes in cognitive function would instead be manifest as intersecting lines, with no apparent differences initially in latencies but with a slower rate of decline in latency or errors over time (Figure 1).

One mechanism to separate learning effects from nonspecific behavioral influences relies on a paradigm such as the multiple schedule of repeated learning and performance (97). This comprises two different behavioral components that alternate over the course of a behavioral test session, with each

component associated with a different environmental stimulus. The active environmental stimulus provides information to the subject about which component is currently operative. In the repeated-learning component the subject is required to learn a sequence of responses, and this sequence changes with each test session in an unpredictable way. This allows the generation of a learning curve during each session. The performance component requires the execution of a sequence of responses of the same length as that in the repeated-learning component, but which has already been learned and remains the same over the course of the experiment. It also requires the same motor, sensory, and motivational capabilities as does behavior in the repeated-learning component but does not require learning per se as long as the task can be learned initially by the subject treated during development. Thus, a true deficit in learning under this schedule would be manifest as a decrease in accuracy in the learning but not in the performance component of the schedule. Concurrent decreases in accuracy in the performance component would be indicative of nonspecific changes in behavior, whether sensory, motor, or motivational, that indirectly contributed to any decreases in the learning component. Figure 2 presents an example of a selective effect on learning following chronic low-level post-weaning lead exposure of rats (97), as indicated by decreases in accuracy in the repeated-learning component and the absence of any such changes in the performance component. Validation of this paradigm in the laboratory requires that the investigator be able to demonstrate that acquisition does indeed occur in the repeated-learning component, i.e., that an increase in accuracy over the course of this component from chance levels can be shown. This approach also has applicability across species ranging from the mouse to the human (94,97–99).

Like the water maze used to measure learning (or short-term memory), simple approaches to measurement of memory such as the frequently employed passive avoidance paradigm also present difficulties of interpretation. This technique relies on the ability of subjects to remember in which compartment of a two-compartment chamber they had previously received shock; the longer it takes for them to re-enter that compartment, the greater the attributed memory. However, difficulties in sensory processing may render the environmental stimuli that dissociate the shocked from the nonshocked compartments as less distinct, thus causing premature re-entries. In cases where the shock training occurs after experimental treatments in between-groups designs, the treatment itself may produce differences in shock sensitivity

that are not apparent in any way but that can influence the subsequent avoidance of the shocked compartment. This can also be checked with a shock titration curve (100).

Paradigms that explicitly control for such alternative explanations include delayed matching-to-sample, which can also be used across species. Memory paradigms typically measure accuracy of remembering following various delay intervals. Increasing delays are associated with increasing difficulty in remembering and thus increases in errors (decreases in accuracy), resulting in a typical delay function (Figure 3, control). To determine the extent to which any alteration in the delay function in response to a treatment is caused by memory impairment rather than changes in other behavioral processes, it is critical to include a no-delay condition (0-sec delay). In this trial no delay is imposed before the subject is asked to match two stimuli, and thus no remembering is required. If deficits in accuracy are observed under these conditions (see nonspecific effect curve in Figure 3), they cannot be ascribed to memory impairments and would suggest that treatment-related decreases in accuracy are non-mnemonic resulting from nonspecific behavioral influences. A true deficit in memory would be reflected instead in a curve in which there were no impairments of accuracy at the 0-sec delay, and increasing delay values would be associated with an increasing decline in accuracy relative to control (Figure 3, “specific” curve). Figure 4 shows a delay function for children 10–11 years of age using the same behavioral test administered from a computerized touch-screen apparatus (101). Such paradigms require the incorporation of delay values that ultimately result in chance levels of accuracy for the species being tested. Using delay values that are too short, and thus do not produce any substantive decline in accuracy, will

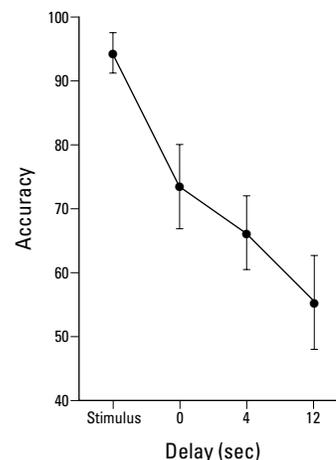


Figure 4. A delay function relating changes in accuracy to delay value (seconds) derived from a sample of 10 normal children 10–12 years of age using the Cantab version of the delayed match to sample procedure (101).

render the paradigm potentially insensitive to the detection of a treatment-related decline.

Research needs. Several research needs merit particular mention. First is the need for additional paradigms for testing cognitive function that can be used earlier in development. The question of long-term adverse consequences usually results in testing of experimental animals in adulthood, but in human populations, tests are used earlier in development to evaluate the ontogeny of such effects. An additional need is for simple assays of learning and memory with adequate sensitivity but that could be used in the context of screening assessments and thus trained more rapidly than more sophisticated procedures such as the multiple schedule of repeated learning and performance and delayed match-to-sample. Finally, a more systematic and refined understanding of attention and its various components will be required to understand its component parts and their underlying anatomic and neurochemical substrates, and how these aspects of attention may be differentially affected by exposures to various toxicants.

Social Behavior

Large portions of the behavioral repertoire of most species are devoted to relations with conspecifics. Aggressive, affiliative, mating, play, and parental behaviors are examples of this category and are among the phenomena most studied by ethologists, biopsychologists, and other life scientists. These behaviors tend to receive less attention from toxicology than assessments of individual behaviors, in part because in the typical laboratory environment rodents—the predominant test species—are not given many opportunities for social interactions. Additionally, measurements of social behaviors are less standardized than, for example, motor activity, and are not as easily incorporated into batteries of screening tests. In addition, because most social behaviors must first be interpreted to be quantified by counts or ratings of defined actions and may sometimes be difficult to automate, they often require trained observers. They may also require the observer to record the responses of two or more animals concurrently, which is another complicating factor.

Social behaviors may be destined to attract more attention from neurobehavioral toxicology because of the types of questions recently aroused by endocrine disrupters. Conspecific behaviors in adults such as mating and aggression are linked directly to prevailing hormonal mechanisms and states, as are behaviors of somewhat greater subtlety such as birdsong patterns and the ordering of dominance hierarchies.

Social behaviors, moreover, are not the exclusive province of hormonally active

agents. They are also modified by many other classes of developmental neurotoxicants that may act either through neuroendocrine mechanisms or by directly influencing brain morphology or neurochemistry. Prenatal alcohol exposure, for example, can impair copulatory behavior in male rats. Prenatal lead exposure intensifies aggressive behaviors in hamsters, as measured by the response to intruders (102). Lead is also a recognized reproductive toxicant. Does this effect represent actions on neuroendocrine status? Many therapeutic agents administered prenatally, such as the benzodiazepine oxazepam, can modify subsequent social behaviors of the offspring such as maternal care (103). Aggressive and defensive behaviors are accompanied by large changes in selected brain dopamine, serotonin, and γ -aminobutyric acid systems (104). Maternal behaviors, aggressive behaviors, and sexual behaviors are among the most promising candidates for social behavior measures in developmental neurotoxicology.

Maternal behavior. Altered endocrine status during fetal development can modify many postnatal behaviors, but little information is available on how prenatal treatment affects maternal behavior in female offspring. If prenatal exposures interfere with endocrine system development, the consequences could appear as abnormalities in maternal behaviors, which are synchronized with a series of hormonal changes that act on the reproductive tract, the mammary gland, and the central nervous system. The immediate hormonal events for maternal behavior occur during pregnancy and also around parturition and lactation, when maternal behavior is fully initiated. Precursors of the full repertoire of maternal behaviors begin during gestation. If pregnant rats are tested for maternal behavior by presenting them with test pups, they show a gradual increase in such behaviors as parturition approaches (nursing posture, licking and retrieving, nest building). After parturition, maternal behavior is maintained essentially by stimulation from the young, such as suckling. A variety of behaviors may be scored in studies of maternal behavior (105,106). These include retrieval of displaced pups, nest building, nursing and licking, and attacks against intruders (107). Instrumental techniques have been reported by Vernotica et al. (108) and Lee et al. (109) that could serve as more automated methods.

Aggressive or attack behaviors. Aggression is a label applied to common responses in many species to invasions of territory, in contesting for mates, in exercising dominance, and even in play behavior. It is among the most frequent social behaviors displayed by animals, including common laboratory species (110). Aggressive behaviors, which consist of several components, including

attack, defensive, and submissive responses, can be modified by many drugs and have been linked to specific neurotransmitter systems (104). In rats, for example, an intruder typically responds to threat postures or attacks by the resident by adopting defensive postures, while the resident may follow threat postures by leaping and biting. Normally, aggressive behaviors in laboratory and house mice are both organized and maintained by testosterone. For the full expression of such behaviors to occur, androgens must be present both during brain development and subsequently. A scoring system has been used to record these and other behaviors on the part of the resident and the intruder (111). Examples of prenatal chemical exposure linked to adult aggressive behaviors are given in Palanza et al. (112) and Fiore et al. (113).

Mating behaviors. Copulatory mechanics are only a minor feature of male sexual behavior, which is driven and organized by the central nervous system. Female sexual behavior is also predominantly dependent on the central nervous system. For these reasons, any plan to study mating behaviors should include situations designed to reveal their behavioral and especially motivational complexities.

Various measures of receptivity that describe female motivation and the reinforcing potency of sex and that simulate the conditions of sexual behavior in natural settings have been devised. For example, a two-compartment test apparatus has been developed in which only the female is able to move from one compartment to the other because of her smaller size (114). A similar approach makes use of a bilevel chamber (115). Such chambers consist of two levels connected by a set of ramps. Because females can run from level to level, the males are forced to follow to attain copulation. In the standard assessment of copulatory function, males are generally provided with a primed female, most often one that has been ovariectomized and then acutely treated with a combination of estradiol and progesterone to induce receptivity. Observers typically record measures of copulatory performance. Copulatory performance in male rats provided an index of interference with gonadal development produced by gestational exposure to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin in a study by Mably et al. (116).

Motivational and incentive measures might, in fact, prove more sensitive to developmental toxicants than scores based simply on the isolated sex act itself, because experimental data show the breadth and complexity of anatomic, neurochemical, and neuroendocrine influences governing sexual motivation (117). Amstislavsky and co-workers (118) relied on a simple technique with mice. They treated pregnant mice with methoxychlor,

then assessed sexual motivation in male offspring by using a cage with a plastic partition behind which they had placed an estrous female. In addition, the bilevel chamber described above has also been used to explore sexual motivation in males (119).

Research needs. Social or conspecific behaviors have received relatively little attention from neurotoxicologists. Naturalistic studies of behavior, the discipline of ethology, have not proven a popular area of neurotoxicity research (120). Now, with the expanding interest in endocrine disruption as an index of toxicity, the appearance of reports linking lead exposure to aggression (121), and public concern that environmental chemicals may be responsible for some antisocial behaviors, the situation is primed for a new look.

If the types of social behaviors described in this review are to be integrated into screening batteries, they must display the attributes common to most of the tests now incorporated into such protocols. Discussed below are three interconnected issues would have to be resolved.

Reliability among observers. Extensive training is required to ensure that different observers in the same laboratory agree on scoring. Attaining agreement among observers demands attention to precise definitions and practice, even for simpler functional observation batteries. Few reports include measures of interobserver reliability or training procedures. Agreement among laboratories must be achieved if social behaviors are to serve as useful end points.

Expertise. Most of the research conducted on social behaviors originates in academic settings, where investigators strive for originality in technique. It would be rare for a group of experts in maternal behavior, for example, to agree on common definitions and approaches so that data from different laboratories can be compared.

Time commitments. It would be difficult to include end points requiring a large investment of investigator or even technician time in screening batteries, even for tier II assessments. Automation is not yet a common feature of social behavior research, although some of the methods described here either have been adapted for it or can be converted without extensive modification.

A number of methods suitable for use as the basis for creation of more efficient techniques for measuring social behaviors have been described in this report. Little standardization has been accomplished, compared with accepted techniques such as functional observation batteries, schedule-controlled operant behavior, and motor activity. Neurobehavioral toxicologists should be in the vanguard of an effort to devise new techniques and to perfect older ones.

Autonomic and Thermoregulatory Function

The autonomic nervous system (ANS) controls the function of a wide variety of organ systems, including the respiratory, cardiovascular, and genitourinary systems. Physiologists and pharmacologists have developed many sophisticated methods to measure the function of these organ systems. Measurement of ANS function has not been a priority for either developmental neurotoxicologists or DNT testing guidelines. In part, this lack of attention may reflect the relative paucity of chemicals that damage the ANS in rats (122). In addition, it is well recognized that the ANS controls vital functions, so that damage to the ANS should significantly compromise general health and/or reproductive capacity. Thus, children with inherited or acquired dysautonomia evince multiorgan disturbances in critical body functions and do not thrive (123). In general, the signs of ANS toxicity are quite obvious. For example, clinical conditions that disrupt the innervation of the bowel are expressed as hyper- or hypomotility states and are reflected by colic, abdominal distension, constipation, or diarrhea. Body weight loss is a frequent correlate (124).

Thermoregulation is accomplished through a network of peripheral and central thermoreceptors and effectors that include somatic (e.g., moving to a warmer or colder location), endocrine, and autonomic (e.g., peripheral vasodilation or vasoconstriction) components (125,126). The function of many components of the thermoregulatory system can be measured in rats using a variety of established test methods. Like autonomic function (see above), thermoregulation has been the purview of physiologists, pharmacologists, and neuroscientists and has not been a priority for either developmental neurotoxicologists or DNT testing guidelines. This inattention may be unfortunate, because thermoregulatory responses to neurotoxicants are an important component of the reaction of adult rats to neurotoxicants (125–127). For example, hyperthermia is an important part of the pathophysiology of the neurotoxic effect of methamphetamine on dopamine-containing nerve terminals in the corpus striatum of the rat (128). Moreover, Gordon and colleagues have shown that perinatal exposure to dioxin can produce long-lasting changes in autonomic and behavioral thermoregulation (129,130).

Biologic Rhythms

Classes of behavior that exhibit biologic rhythms include feeding, drinking, sleeping, motor activity, and mating (131). The cycle associated with each of these behaviors represents a potential tool that could be examined for potential effects of chemical treatment.

For example, chemicals may disrupt or alter the diurnal pattern of locomotor activity (132,133), and there are chemicals that elicit either a more pronounced or diminished diurnal pattern of locomotor activity in rats. In addition, diurnal patterns of ingestion exhibit changes not evident by measures of the total amount ingested (134). Triethyltin, for example, alters the diurnal pattern of water ingestion but not total daily consumption (132), whereas trimethyltin increases total water consumption while the diurnal pattern is largely preserved. Although the value of such approaches is illustrated by these examples, this approach to neurotoxicity assessment remains relatively unexplored.

Considered more relevant to the present discussion is how biologic rhythms or cycles can affect behavioral test results by introducing additional variability and complicating the interpretation of behavioral test results. For example, the level of activity exhibited by animals over the course of a day is one of the most apparent and well-established behavioral manifestations of circadian rhythms. Within an 8-hr workday, levels of horizontal and vertical activity vary by as much as 20–30% (135). If not adequately controlled, this can contribute substantially to variability in measures of motor activity. Hormonal cycles may also contribute to variability. Levels of activity in the running wheel are 3–10 times higher for female rats in estrous than levels during diestrous (136). Diurnal factors have also been shown to affect the ability of tests to detect the effects of certain chemical treatments (137,138). Thus, circadian rhythms represent a significant source of variability for behavioral test results. If left unmanaged, statistical power will be reduced, increasing the probability of a type II error. One approach used to compensate for this is to increase the number of animals in each dose group. However, the potential gain associated with this approach may not be realized if appropriate precautions are not taken, as additional time will be required to test those animals. Thus, to the extent possible, appropriate measures should be incorporated into the study design, such as including representatives from each dose group in each set of animals tested at one time, and testing those animals over more days rather than extending hours on a given test day.

Conclusions

Careful consideration of a number of experimental design issues is the key to the success of a study using behavioral methods to examine DNT. Identifying clear study goals and objectives is paramount in designing a strong study. Study goals and objectives are a guide in selection of the methods used, the appropriate animal model, and the equipment

needed. The study goals can be used to identify the behavioral test methods by guiding the evaluation of the sensitivity versus selectivity required to answer the scientific question at hand. Method selection also rests on consideration of available resources, including equipment, funds, and personnel. An understanding of the inherent variability in the methods selected can and should be used to determine the number of animals required to detect an effect of concern. Variability can arise from a number of different sources in tests of sensory function, and this is particularly true in studying effects on developing animals. Averaging response data across ages in developing animals could increase variability unnecessarily, as the response may change drastically in a very short time because of increases in body weight and increased sensitivity of the developing sensory system. Appropriate statistical analyses are vital for defensible interpretation of behavioral data. Repeated measures are often used in DNT testing and must be treated as such in the statistical analysis. When using behavioral methods in toxicity testing, it is important that normative data be developed for a method demonstrating that the test method as used can detect and characterize the effects of varying the magnitude of these properties. Positive and negative control data are needed to support the validation of the method and aid in data interpretation.

Proper expertise is required to design, conduct, and interpret a study of developmental neurotoxicity using behavioral methods. Training in experimental psychology or psychopharmacology and in statistics provides a background important for design and interpretation of these studies. Those who conduct the tests should be trained in proper performance of each behavioral test and should have a good understanding of potential confounders.

Behavioral methods are used to detect and characterize developmental neurotoxic effects on sensory cognitive and motor system functions. The major sensory systems of concern in toxicology include visual, auditory, olfactory, nociceptive (pain and other noxious stimuli), somatosensory, and vestibular. There are a number of stimulus properties shared by all sensory systems, including intensity, frequency, duration, and location in space. One tenant of sensory function testing is that it never yields a direct measurement of sensation; instead, a change in sensory function is inferred based on the observed change in the motor response evaluated. A better understanding of the relationship between the simple sensory system tests currently used in DNT studies and any underlying changes in the structure and/or function of the sensory system will advance our ability to identify sensory

system toxicity. Behavioral toxicologists can continue to learn from recent advances in neurobiology and genomics that may allow an increased understanding of the physiologic and structural bases for sensory function.

Behavioral tests of motor dysfunction in animals include those used to detect spontaneous movement disorders such as changes in gait, tremors, and myoclonus, and those used to detect changes in induced movement such as reflexes, reactions, and movements under operant control. Tests of motor function include observation of locomotion, measurement of locomotor activity, and tests of reflexes and reactions. Patterns of changes in several reflexes and reactions are generally characterized by neurologic examination. There is a need to develop and validate technology and procedures that measure motor function objectively and sensitively, yet are flexible enough to be used with large numbers of animals. Interpretation of motor function tests must take into account the limitations of the test equipment as well as any potential biologic confounders.

Cognitive function is thought to encompass learning, memory, and attention processes. Assessment of cognitive function is a critical component of a DNT assessment to address concerns over potential long-term consequences of exposures to toxicants during brain development. In many simpler tests of learning and memory, changes in sensory function, motor behavior, and/or motivation may indirectly influence the dependent measures used, and therefore change behavior. Simple paradigms generally do not include control procedures for assessing these possibilities. Reliance on relatively complex approaches to answer questions about potential cognitive deficits could be useful, even in screening, to determine whether a potential learning/memory impairment might be attributable to deficits in other areas of nervous system function.

Social behaviors, such as aggressive, affiliative, mating, play, and parental behaviors tend to receive less attention from toxicologists than individual behaviors. Social behaviors may be modified by developmental neurotoxicants, including hormonally active agents that may act through neuroendocrine mechanisms or by directly influencing brain morphology or neurochemistry. Techniques used to measure social behaviors are less standardized than individual behaviors. Because most social behaviors have to be interpreted to be quantified, the tests are difficult to automate and trained observers are often required to perform the tests. If the types of social behaviors described in this review are to be integrated into screening batteries, they will have to display the attributes common to most of the tests now incorporated into such protocols.

Measurement of ANS function has not been a priority for developmental neurotoxicologists, though a number of sophisticated tests developed by physiologists, pharmacologists, and neuroscientists could be adopted. This inattention may be unfortunate because thermoregulatory responses to neurotoxicants are an important component of the reaction of mammals to neurotoxicants. Classes of behavior that exhibit biologic rhythms include feeding, drinking, sleeping, motor activity, and mating. The cycle associated with each of these behaviors represents a tool that could be examined to identify effects of chemical treatment.

Behavioral testing methods to measure sensory, motor, and cognitive function are well developed, but there is room for improvement in study design, conduct, analysis, and interpretation. Tests to characterize effects of developmental neurotoxicants on social behavior, the ANS, thermoregulation, and circadian rhythms are presently underutilized.

REFERENCES AND NOTES

- Aminoff MJ, Ross A. *Electrodiagnosis*. In: *Clinical Neurology*. New York:Churchill Livingstone, 1999.
- Lotti M. Organophosphorus compounds. In: *Experimental and Clinical Neurotoxicology*, 2nd ed (Spencer PS, Schaumburg HH, eds). New York:Oxford University Press, 2000;897–925.
- Ross JF, Mattsson JL, Fix AS. Expanded clinical observations in toxicity studies: historical perspectives and contemporary issues. *Regul Toxicol Pharmacol* 28:17–26 (1998).
- Moser VC, Ross JF. *Training Video and Reference Manual for a Functional Observational Battery*. Research Triangle Park, NC:U.S. Environmental Protection Agency, 1996.
- Abbey H, Howard E. Statistical procedures in developmental studies on a species with multiple offspring. *Dev Psychobiol* 6:329–335 (1973).
- Cox C. Statistical issues for animal studies in developmental neurotoxicity. In: *Neurobehavioral Toxicity: Analysis Interpretation* (Weiss B, O'Donoghue J, eds). New York:Raven Press, 1994;93–101.
- Holson RR, Pearce B. Principles and pitfalls in the analysis of prenatal treatment effects in multiparous species. *Neurotoxicol Teratol* 14:221–228 (1992).
- Reilly EP, Meyer LJ. Consideration for the design, implementation and interpretation of animal models of fetal alcohol effects. *Neurotoxicol Teratol* 6:97–101 (1984).
- U.S. EPA. *Developmental Neurotoxicity Study*. OPPTS 870.6300. EPA 712-C-98-239. Washington, DC:U.S. Environmental Protection Agency, 1998.
- Levin ED, Simon BB, Schmechel DE, Glasgow HB Jr, Deamer-Melia NJ, Burkholder JM, Moser VC, Jensen K, Harry GJ. Pfisteria toxin and learning performance. *Neurotoxicol Teratol* 21(3):215–221 (1999).
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ:Lawrence Erlbaum, 1987.
- Muller KE, Benignus VA. Increasing scientific power with statistical power. *Neurotoxicol Teratol* 14:211–219 (1992).
- Crofton KM, Howard JL, Moser VC, Gill MW, Reiter LW, Tilson H, MacPhail RC. Interlaboratory comparison of motor activity experiments: implications for neurotoxicological assessments. *Neurotoxicol Teratol* 13:599–609 (1991).
- Wise LD, Allen HL, Hoe CMVDR, Gerson RJ. Developmental neurotoxicity evaluation of the avermectin pesticide, emamectin benzoate, in Sprague-Dawley rats. *Neurotoxicol Teratol* 19(4):315–326 (1997).
- Fox DA. Sensory system alterations following occupational exposure to chemicals. In: *Occupational Neurotoxicology* (Costa LG, Manzo L, eds). Boca Raton, FL:CRC Press, 1998;169–184.
- Morata TC, Dunn DD. *Occupational Medicine: State of Art Reviews*. Philadelphia:Henley and Belfus, 1995.
- Anger WK, Johnson BL. Chemical affecting behavior. *Neurotox Ind Comm Chem* 1:51–148 (1985).

18. Crofton KM, Sheets LP. Evaluation of sensory system function using reflex modification of the startle response. *J Am Coll Toxicol* 8:199–211 (1989).
19. Maurissen JPJ. Neurobehavioral methods for the evaluation of sensory functions. In: *Neurotoxicology: Approaches and Methods* (Chang L, Slikker W, eds). New York:Academic Press, 1995:239–264.
20. Willott JF. Anatomic and physiologic aging: a behavioral neuroscience perspective. *J Am Acad Audiol* 7:141–151 (1996).
21. Henley CM, Rybak LP. Developmental ototoxicity. *Otolaryngol Clin North Am* 26:857–871 (1993).
22. Henley CM, Rybak LP. Ototoxicity in developing mammals. *Brain Res Rev* 20:68–90 (1995).
23. Mattsson JL. Ototoxicity: an argument for evaluation of the cochlea in safety testing in animals. *Toxicol Pathol* 28:137–141 (2000).
24. Moody DB, Stebbins WC. Detection of the effect of toxic substances on the auditory system by behavioral methods. In: *Neurotoxicology* (Mitchell CL, ed). New York:Raven Press, 1982:109–132.
25. Stebbins WC. Concerning the need for more sophisticated animal models in sensory behavioral toxicology. *Environ Health Perspect* 44:77–85 (1982).
26. Stebbins WC, ed. *Animal Psychophysics: The Design and Conduct of Sensory Experiments*. New York:Appleton Century-Crofts, 1970.
27. Pryor GT, Bingham LR, Dickinson J, Rebert CS, Howd RA. Importance of schedule of exposure to hexane in causing neurotoxicity. *Neurobehav Toxicol Teratol* 4:71–78 (1982).
28. Merigan WH, Barkdoll E, Maurissen JP. Acrylamide-induced visual impairment in primates. *Toxicol Appl Pharmacol* 62:342–345 (2000).
29. Muller M. Developmental changes of frequency representation in the rat cochlea. *Hear Res* 56:1–7 (1991).
30. Rubel EW. Strategies and problems for future studies of auditory development. *Acta Otolaryngol Suppl* 421:114–128 (1985).
31. Fukumura M, Sithi N, Voorhes CV. Effects of the neurotoxin 3,3'-iminodipropionitrile on acoustic startle and locomotor activity in rats: a comparison of functional observational and automated startle assessments. *Neurotoxicol Teratol* 20:203–211 (1998).
32. Lochry EA. Concurrent use of behavioral/functional testing in existing reproductive and developmental toxicity screens: practical considerations. *J Am Coll Toxicol* 6:433–439 (1987).
33. Merigan WH. Effects of toxicants on visual systems. *Neurobehav Toxicol* 1(suppl 1):15–23 (1979).
34. Wood RW. Behavioral evaluation of sensory irritation evoked by ammonia. *Appl Pharmacol* 50:157–162 (1979).
35. Maurissen JPJ, Weiss B, Davis HT. Somatosensory thresholds in monkeys exposed to acrylamide. *Toxicol Appl Pharmacol* 71:266–279 (1983).
36. Kelly JB, Masterton B. Auditory sensitivity of the albino rat. *J Comp Physiol Psychol* 91:930–936 (1977).
37. Crofton KM, Ding D, Padich R, Taylor M, Henderson D. Hearing loss following developmental exposure to polychlorinated biphenyls: a cochlear site of action. *Hear Res* 144:196–204 (2000).
38. Young JS, Fechter LD. Reflex inhibition procedures for animal audiometry: a technique for assessing ototoxicity. *J Acoust Soc Am* 73:1686–1693 (1983).
39. Ison JR. Reflex modification as an objective test for sensory processing following toxicant exposure. *Neurobehav Toxicol Teratol* 6:437–445 (1984).
40. Evans HL. Assessment of vision in behavioral toxicology. In: *Nervous System Toxicology* (Mitchell CL, ed). New York:Raven Press, 1982:81–108.
41. Fay RR. *Hearing in Vertebrates: A Psychophysics Databook*. Winnetka, IL:Hill-Fay Associates, 1988.
42. Gonda Z, Miklosi A, Lehotzky K. The effect of social learning on a conditioned avoidance response of rats treated prenatally with aluminum lactate. *Neurotoxicol Teratol* 19:59–63 (1997).
43. Glowa JR, Hansen CT. Differences in response to an acoustic startle stimulus among forty-six rat strains. *Behav Genet* 24:79–84 (1994).
44. Sataloff RT, Sataloff J. *Hearing Loss*. New York:Marcel Dekker, 1993.
45. Tilson HA, Burne TA. Effects of triethyl tin on pain reactivity and neuromotor function of rats. *J Toxicol Environ Health* 8:317–324 (1981).
46. Boyes WK, Jenkins DE, Dyer RS. Chlordimeform produces contrast-dependent changes in visual evoked potentials of hooded rats. *Exp Neurol* 89:391–407 (1985).
47. Ross JF. Tier 1 neurological assessment in regulated animal safety studies. In: *Neurotoxicology Handbook, Vol 3* (Massaro EJ, ed). Totowa, NJ:Humana Press, in press.
48. Gallistel CR. Animal cognition: the representation of space, time and number. *Annu Rev Psychol* 40:155–189 (1989).
49. Tchermichowski O, Golani I. A phase plane representation of rat exploratory behavior. *J Neurosci Methods* 62(1–2):21–27 (1995).
50. Mullenix PJ. The computer pattern recognition system for study of spontaneous behavior of rats. In: *Motor Activity and Movement Disorders: Research Issues and Applications* (Sanberg PR, Ossenkopp KP, Kavaliers, eds). Totowa, NJ:Humana Press, 1996:141–174.
51. Paulus MP, Geyer MA, Gold LH, Mandell AJ. Application of entropy measures derived from the ergodic theory of dynamical systems to rat locomotor behavior. *Proc Natl Acad Sci USA* 87:723–727 (1990).
52. Dai H, Carey RJ. A new method to quantify behavioral attention to a stimulus object in a modified open-field. *J Neurosci Methods* 53:29–34 (1994).
53. File S. Pharmacological manipulations of responses to novelty and their habituation. In: *Theory in Psychopharmacology* (Cooper SJ, ed). New York:Academic Press, 1981:197–232.
54. MacPhail RC, Peele DB, Crofton KM. Motor activity and screening for neurotoxicity. *J Am Coll Toxicol* 8:117–125 (1989).
55. Maurissen JPJ, Mattsson JL. Critical assessment of motor activity as a screen for neurotoxicity. *Toxicol Ind Health* 5:195–202 (1989).
56. Meyer OA, Tilson HA, Byrd WC, Riley MT. A method for the routine assessment of fore- and hindlimb grip strength of rats and mice. *Neurobehav Toxicol* 1:233–236 (1979).
57. Ross JF, Handley DE, Fix AS, Lawhorn GT, Carr GJ. Quantification of the hindlimb extensor thrust response in rats. *Neurotoxicol Teratol* 19:405–411 (1997).
58. Edwards PM, Parker VH. A simple, sensitive, and objective method for early assessment of acrylamide neuropathy in rats. *Toxicol Appl Pharmacol* 40:589–591 (1977).
59. U.S. EPA. U.S. Environmental Protection Agency Health Effects Guidelines, Neurotoxicity Screening Battery. OPPTS 870.6200. Washington, DC:U.S. Environmental Protection Agency, 1998.
60. Cartmell J, Monn JA, Schoep DD. The metabotropic glutamate 2/3 receptor agonists LY354740 and LY379268 selectively attenuate phenylcyclidine versus *d*-amphetamine motor behaviors in rats. *J Pharmacol Exp Ther* 291:161–170 (1999).
61. Altman J, Sudarshan K. Postnatal development of locomotion in the laboratory rat. *Anim Behav* 23:896–920 (1975).
62. Fox WM. Reflex-ontogeny and behavior development of the mouse. *Anim Behav* 13:234–241 (1965).
63. Stanford JA, Fowler SC. Subchronic effects of clozapine and haloperidol on rats' forelimb force and duration during press-while-licking-task. *Psychopharmacology (Berl)* 130(3):249–253 (1997).
64. Eilam D, Golani I. Home base behavior of rats (*Rattus norvegicus*) exploring a novel environment. *Behav Brain Res* 34(3):199–211 (1989).
65. Tchermichowski O, Golani I. A phase plane representation of rat exploratory behavior. *J Neurosci Methods* 62:21–27 (1995).
66. Archer J. Tests for emotionality in rats and mice: a review. *Anim Behav* 21:205–235 (1973).
67. Robbins TW. A critique of the methods available for the measurement of spontaneous motor activity. In: *Principles of Behavioral Pharmacology*. New York:Plenum Press, 1977.
68. Walsh RN, Cummins RA. The open-field test: a critical review. *Psychol Bull* 83:482–504 (1976).
69. Etienne AS, Maurer R, Seguinot V. Path integration in mammals and its interaction with visual landmarks. *J Exp Biol* 199(Pt 1):201–209 (1996).
70. Whishaw IQ, Brooks BL. Calibrating space: exploration is important for allothetic and idiothetic navigation. *Hippocampus* 9:659–667 (1999).
71. Hine B. Differential open field reactivity in HAS and LAS rats. *Physiol Behav* 57(2):301–306 (1995).
72. Treit D, Fundytus M. Thigmotaxis as a test for anxiolytic activity in rats. *Pharmacol Biochem Behav* 31(4):959–962 (1988).
73. Schwarting RKW, Fornagueria J, Hutson JP. Automated video-image analysis of behavioral asymmetries. In: *Motor Activity and Movement Disorders: Research Issues and Applications*. Totowa, NJ:Humana Press, 1996.
74. Ross JF. EOCs, FOBs and UFOs. Making sense of observational data. *Toxicol Pathol* 28:132–136 (2000).
75. Sanberg PR, Zoloty SA, Willis R, Ticarich CD, Rhoads K, Nagy RP, Mitchell SG, Laforest AR, Jenks JA, Harkabas LJ, et al. Digiscan activity: automated measurement of thigmotactic and stereotypic behavior in rats. *Pharmacol Biochem Behav* 27(3):569–572 (1987).
76. Geyer MA, Russo PV, Masten VL. Multivariate assessment of locomotor behavior: pharmacological and behavioral analyses. *Pharmacol Biochem Behav* 25:277–288 (1986).
77. Kernan WJ Jr, Mullenix PJ, Hopper DL. Time structure analysis of behavioral acts using a computer pattern recognition system. *Pharmacol Biochem Behav* 34(4):863–869 (1989).
78. Paulus MP, Geyer MA. A temporal and spatial scaling hypothesis for the behavioral effects of psychostimulants. *Psychopharmacology (Berl)* 104(1):6–16 (1991).
79. Paulus MP, Callaway CW, Geyer MA. Quantitative assessment of the microstructure of rat behavior. II: Distinctive effects of dopamine releasers and uptake inhibitors. *Psychopharmacology (Berl)* 113:187–198 (1993).
80. Paulus MP, Geyer MA. Quantitative assessment of the microstructure of rat behavior. I: The extension of the scaling hypothesis. *Psychopharmacology (Berl)* 113:177–186 (1993).
81. Young MS, Young CW, Li YC. A combined system for measuring animal motion activities. *J Neurosci Methods* 95:55–63 (2000).
82. Cory-Slechta DA. Bridging experimental animal and human behavioral toxicology studies. In: *Behavioral Measures of Neurotoxicity* (Russel RW, Flattau PE, Pope AM, eds). Washington, DC:National Academy Press, 1990:137–158.
83. Carson TL, VanGelder GA, Karas GC, Buck WB. Slowed learning in lambs prenatally exposed to lead. *Arch Environ Health* 29:154–156 (1974).
84. Wood R, Rees D, Laties V. Behavioral effects of toluene are modulated by stimulus control. *Toxicol Appl Pharmacol* 68:462–472 (1983).
85. Neitz J, Jacobs G. Reexamination of spectral mechanisms in the rat (*Rattus norvegicus*). *J Comp Psychol* 100(1):21–29 (1986).
86. Hastings L. Sensory neurotoxicology: use of the olfactory system in the assessment of toxicity. *Neurotoxicol Teratol* 12:455–459 (1990).
87. Chen JS, Amsel A. Retention under changed-reward conditions of persistence learned by infant rats. *Dev Psychobiol* 13:469–480 (1980).
88. Dominguez HD, Bocco G, Chotro MG, Spear NE, Molina JC. Operant responding controlled by milk or milk contaminated with alcohol as positive reinforcers in infant rats. *Pharmacol Biochem Behav* 44:403–409 (1993).
89. Lobaugh NJ, Greene PL, Grant M, Nick T, Amsel A. Patterned (single) alternation in infant rats after combined or separate lesions of hippocampus and amygdala. *Behav Neurosci* 103:1159–1167 (1989).
90. Moyer TB, Vanderryn J. Physostigmine accelerates the development of associative memory processes in the infant rat. *Psychopharmacology* 95:401–406 (1988).
91. Hughes C, Plumet MH, Leboyer M. Towards a cognitive phenotype for autism: increased prevalence of executive dysfunction and superior spatial span amongst siblings of children with autism. *J Child Psychol Psychiatry* 40:705–718 (1999).
92. Kempton S, Vance A, Maruff P, Luk E, Costin J, Pantelis C. Executive function and attention deficit hyperactivity disorder: stimulant medication and better executive function performance in children. *Psychol Med* 29:527–538 (1999).
93. Luciana M, Nelson CA. The functional emergence of prefrontally-guided working memory systems in four-to-eight year old children. *Neuropsychologia* 36:273–293 (1998).
94. Brooks AJ, Cory-Slechta DA, Federoff HJ. Repeated acquisition and performance chamber for mice: a method to assess spatial learning and memory. *Neurobiol Learn Mem* 74(3):241–258 (2000).
95. Kim C, Nakai K, Kasanuma Y, Satoh H. Comparison of neuro-behavioral changes in the three inbred strains of mice prenatally exposed to methylmercury. *Neurotoxicol Teratol* 22(3):397–403 (2000).
96. Kuhlmann AC, McGlothlin JL, Guilarte TR. Developmental lead exposure causes spatial learning deficits in adult rats. *Neurosci Lett* 233:101–104 (1997).
97. Cohn J, Cox C, Cory-Slechta D. The effects of lead exposure on learning in a multiple repeated acquisition and performance schedule. *Neurotoxicology* 14:329–346 (1993).
98. Kelly TH, Foltin RW, Fischman MW. The effects of repeated amphetamine exposure on multiple measures of human behavior. *Pharmacol Biochem Behav* 38:417–426 (1991).
99. Moerschbaecher JM, Thompson DM, Winsauer PJ. Effects of opioids and phenylcyclidine in combination with naltrexone on the acquisition and performance of response sequences in monkeys. *Pharmacol Biochem Behav* 22:1061–1069 (1985).
100. Peele DB, Allison SD, Crofton KM. Learning and memory deficits in rats following exposure to 3,3'-iminodipropionitrile. *Toxicol Appl Pharmacol* 105:321–332 (1990).
101. Cory-Slechta DA. Unpublished observations (2000).
102. Delville Y. Exposure to lead during development alters aggressive behavior in golden hamsters. *Neurotoxicol Teratol* 21:445–449 (1999).

103. Petrucci S, Chiarotti F, Alleva E, Laviola G. Limited changes of mouse maternal care after prenatal oxazepam: dissociation from pup-related stimulus perception. *Psychopharmacology (Berl)* 122(1):58–65 (1995).
104. Miczek KA, Weerts E, Haney M, Tidey J. Neurobiological mechanisms controlling aggression: preclinical developments for pharmacotherapeutic interventions. *Neurosci Biobehav Rev* 18:97–110 (1994).
105. Lonstein JS, Wagner CK, De Vries GJ. Comparison of the “nursing” and other parental behaviors of nulliparous and lactating female rats. *Horm Behav* 36:242–251 (1999).
106. Stern JM. Maternal behavior: sensory, hormonal and neural determinants. In: *Psychoneuroendocrinology* (Brush FR, Levine S, eds). New York:Academic Press, 1989:105–226.
107. Lonstein JS, Stern JM. Site and behavioral specificity of periaqueductal gray lesions on postpartum sexual, maternal, and aggressive behaviors in rats. *Brain Res* 804(1):21–35 (1998).
108. Vernotica EM, Rosenblatt JS, Morrell JI. Microinfusion of cocaine into the medial preoptic area or nucleus accumbens transiently impairs maternal behavior in the rat. *Behav Neurosci* 113:377–390 (1999).
109. Lee A, Li M, Watchus J, Fleming AS. Neuroanatomical basis of maternal memory in postpartum rats: selective role for the nucleus accumbens. *Behav Neurosci* 113(3):523–538 (2000).
110. Scott JP. *Animal Behavior*. Chicago:University of Chicago Press, 1972.
111. Miczek KA, Winslow JT, DeBold JF. Heightened aggressive behavior by animals interacting with alcohol-treated conspecifics: studies with mice, rats, and squirrel monkeys. *Pharmacol Biochem Behav* 20:349–353 (1984).
112. Palanza P, Parmigiani S, Liu H, vom Saal FS. Prenatal exposure to low doses of the estrogenic chemicals diethylstilbestrol and *o,p'*-DDT alters aggressive behavior of male and female house mice. *Pharmacol Biochem Behav* 64(4):665–672 (1999).
113. Fiore M, Petrucci S, Dell’Omo G, Alleva E. Prenatal sulfur dioxide exposure induces changes in the behavior of adult male agonistic encounters. *Neurotoxicol Teratol* 20:543–548 (1998).
114. Paredes RG, Vazquez B. What do female rats like about sex? Paced mating. *Behav Brain Res* 105:117–127 (1999).
115. Pfau JG, Smith WJ, Coopersmith CB. Appetitive and consummatory sexual behaviors of female rats in bilevel chambers. I: A correlational and factor analysis and the effects of ovarian hormones. *Horm Behav* 35:224–240 (1999).
116. Mably TA, Moore RW, Goy RW, Peterson RE. In utero and lactational exposure of male rats to 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. 2. Effects on sexual behavior and the regulation of luteinizing hormone secretion in adulthood. *Toxicol Appl Pharmacol* 114:108–117 (1992).
117. Hull EM, Lorrain DS, Du J, Matuszewich L, Lumley LA. Hormone-neurotransmitter interactions in the control of sexual behavior. *Behav Brain Res* 105:105–106 (1999).
118. Amstislavsky SY, Amstislavskaya TG, Eroschenko VP. Methoxychlor given in the periimplantation period blocks sexual arousal in male mice. *Reprod Toxicol* 13:405–411 (1999).
119. Mendelson SD, Pfau JG. Level searching: a new assay of sexual motivation in the male rat. *Physiol Behav* 45:337–341 (1989).
120. Cohn J, MacPhail RC. Ethological and experimental approaches to behavior analysis: implications for ecotoxicology. *Environ Health Perspect* 104(2):299–305 (1996).
121. Needleman HL, Reiss JA, Tobin MJ, Biesecker GE, Greenhouse JB. Bone lead levels and delinquent behavior. *JAMA* 275:363–369 (1996).
122. Spencer PS, Schaumburg HH, eds. *Experimental and Clinical Neurotoxicology*. New York:Oxford University Press, 2000.
123. Smith SA. Disorders of the autonomic nervous system. In: *The Practice of Pediatric Neurology* (Swaiman KF, Wright FS, eds). St. Louis, MO:C.V. Mosby, 1975:799–828.
124. Oribe E, Appenzeller O. Autonomic nervous system. In: *Guide to Clinical Neurology* (Mohr JP, Gautier JC, eds). New York:Churchill Livingstone, 1995:521–540.
125. Gordon CJ, Heath JE. Integration and central processing in temperature regulation. *Annu Rev Physiol* 48:595–612 (1986).
126. Gordon CJ. Thermal biology of the laboratory rat. *Physiol Behav* 47:963–991 (1990).
127. Watkinson WP, Gordon CJ. Caveats regarding the use of the laboratory rat as a model for acute toxicological studies: modulation of the toxic response via physiological and behavioral mechanisms. *Toxicology* 81:15–31 (1993).
128. Bowyer JF, Davies DL, Schmued L, Broening HW, Newport GD, Slikker WJ, Holson RR. Further studies of the role of hyperthermia in methamphetamine neurotoxicity. *J Pharmacol Exp Ther* 268:1571–1580 (1994).
129. Gordon CJ, Yang Y, Gray LE Jr. Autonomic and behavioral thermoregulation in golden hamsters exposed perinatally to dioxin. *Toxicol Appl Pharmacol* 137:120–125 (1996).
130. Gordon CJ, Miller DB. Thermoregulation in rats exposed perinatally to dioxin: core temperature stability to altered ambient temperature, behavioral thermoregulation, and febrile response to lipopolysaccharide. *J Toxicol Environ Health* 54:647–662 (1998).
131. Rusak B, Zucker I. Biological rhythms and animal behavior. *Annu Rev Psychol* 27:137–171 (1975).
132. Evans HL. Behaviors in the home cage reveal toxicity: recent findings and proposals for the future. *J Am Coll Toxicol* 8:35–52 (1989).
133. MacPhail RC, Crofton KM, Reiter LW. Use of environmental challenges in behavioral toxicology. *Fed Proc* 42:3196–3200 (1983).
134. Evans HL. Neurotoxicity expressed in naturally occurring behavior. In: *Neurobehavioral Toxicity: Analysis and Interpretation* (Weiss B, O’Donoghue J, eds). New York:Raven Press, 1994:111–135.
135. Crofton KM, MacPhail RC. Reliability of motor activity assessments. In: *Measuring Movement and Locomotion: From Invertebrates to Humans* (Ossenkopp KP, Kavaliers M, Sanberg RG, eds). Austin, TX:Landes, 1996:227–252.
136. Wang GH. The relation between “spontaneous” activity and oestrous cycle in the white rat. *Comp Psychol Monogr* 6:2 (1923).
137. Norton S, Culver B, Mullenix P. Measurement of the effects of drugs on activity of permanent groups of rats. *Psychopharmacol Commun* 1(2):131–138 (1975).
138. Wolfe G, Bousquet W, Schnell R. Circadian variations in response to amphetamine and chlorpromazine in the rat. *Commun Pharmacol* 1:29–37 (1977).