Original article

Assessing behavior and cognition in rodents, nonhuman primates, and humans: where are the limits of translation?

Marius Stephan, MSc; Paul Volkmann; Moritz J. Rossner, PhD

New psychopharmacological treatments are needed for affective and nonaffective psychoses, especially for the associated negative and cognitive symptoms. Earlier developments mostly failed, probably partly because of limitations in behavioral models used for validation. Now, deeper understanding of the genetics underlying disease pathogenesis and progress in genetic engineering will generate many rodent models with increased construct validity. To improve these models' translational value, we need complementary data from nonhuman primates. We also have to improve and streamline behavioral test systems to cope with increased demand. Here, we propose a comprehensive neurocognitive test battery that should overcome the disadvantages of single tests and yield cognitive/behavioral profiles for modeling subsets of patient symptoms. Further, we delineate a concept for classifying disease-relevant cognitive endophenotypes to balance between face and construct validity and clinical diagnostics. In summary, this review discusses new concepts and the limitations and future potential of translational research on cognition in psychiatry.

Introduction

Among the mental disorders, major depressive disorder (MDD), bipolar disorder (BD), and schizophrenia (SZ) —collectively termed affective and nonaffective psychoses—together cause the highest number of years lived with disability worldwide.[1] Large genome-wide association studies (GWASs) have revealed more than 100 genetic risk loci for SZ, 30 for BD, and 44 for MDD, pinpointing hundreds of implicated genes.[2-4] Moreover, these three disorders share close genetic relationships,[5] affect similar brain regions, and have similar brain transcriptome profiles.[6] The success of GWASs and the advent of human induced pluripotent stem cells (hiPSCs) fostered a better understanding of the highly polygenic genetic and cross-disorder architecture of psychiatric diseases. Among the most prominent mechanisms identified are those modulating neuronal gene expression, synapse-to-nucleus Ca^{2+} signaling, synaptogenesis, and synaptic pruning, as well as alterations of glutamatergic and GABAergic signaling that change the excitation-inhibition (E/I) balance. All of these mechanisms are core neurodevelopmental processes that are strongly associated with synaptic plasticity, circuit formation, and, ultimately, higher-order cognitive performance.

So far, the development of new pharmacotherapeutic compounds has not advanced at the same speed as the research described above. Reasons for the slow advance-

Author affiliations: Molecular and Behavioural Neurobiology, Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Germany (Marius Stephan, Paul Volkmann, Moritz J. Rossner); International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany (Marius Stephan). **Address for correspondence:** Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich,, Nussbaumstr. 7, 80336 Munich, Germany. (email: moritz.rossner@med.uni-muenchen.de)

ment include the lack of genetically validated targets and mechanisms and the almost exclusive focus of decades of academic preclinical research and industrial drug discovery attempts on aminergic signaling. First-line medication for SZ, for example, is still mainly restricted to second-generation antipsychotics, such as risperidone,[7] and the mode of action beyond the proposed effects of antipsychotics mediated through the dopamine D2 receptor still remains a mystery.

Although antipsychotics are effective in treating positive symptoms, such as hallucinations, their efficacy in negative and cognitive symptoms is low, leaving patients with a reduced quality of life and impaired cognitive performance. Additionally, about 20% to 30% of patients with SZ are treatment-resistant, stressing the need for better compounds.[8,9] Research on model systems is likely to be an essential tool in the development of new pharmacotherapeutics targeting cognition, although translation of higher-order cognitive processes remains a difficult challenge. To produce valid and reliable results of translational value, we need new genetic models with higher construct validity (ie, that more closely reflect the molecular cause in patients) and new concepts that more reliably assess cognitive performance in those models and have better subdomain-focused face validity (ie, that evaluate defects in a cognitive domain of relevance in patients). In mice, we are now able to generate genetic models that are based on individual sets of validated risk genes derived from large-scale human genetic databases.[10] Moreover, progress in genome engineering technologies, such as CRISPR/Cas, is likely to evolve towards more refined mouse models in which clustered arrays of risk alleles may further increase the construct validity of complex genetic disorders. In parallel, research on nonhuman primates (NHPs) may complement genetic mouse models because the social behaviors of NHPs, and therefore probably also their psychosocial stressors, may better align with those of humans. Indeed, environmental, chemical, and surgical interventions have been applied to generate such disease models.[1-13] To examine genetic and environmental risks, both of which serve as triggers of pathogenesis and are major determinants of therapy outcome, we need to use

> **Models and experimental procedures remain an essential tool for understanding psychiatric disorders and developing new pharmacotherapeutic compounds**

meaningful rodent and NHP models that include prenatal complications, such as early-life trauma and psychosocial stress in adolescence.[14-16] No disease model, however, will ever fully reflect the human situation because: (i) prototypical clinical symptoms, such as hallucinations, cannot be studied; and (ii) psychiatric diseases are nowadays considered to represent a continuum of cross-disorder clinical and neurobiological phenotypes.[16] Therefore, we should rather aim at modeling a subset of specific and accessible endophenotypes, which we refer to as behavioral and/or cognitive subdomains. We think that such a stratification towards defined "subdomain-oriented" animal models may represent a better means for validating compounds targeting novel eg, GWAS-derived mechanisms in the future. An important and challenging task, however, is to cope with the increased demand in characterizing novel models that are likely to be developed to deconvolute risk gene/phenotype relationships. A central goal is to identify technically rather simple, robust, and valid translational behavioral tests for a broad spectrum of cognitive capabilities and to organize these into a pipeline for rapid and comprehensive screening in rodent and NHP models. Therefore, in the following sections we will review tests that have been developed to assess behavior and higher-order cognition in rodent models, NHPs, and humans. We will focus in particular on subdomain translatability and practical considerations with respect to handling/training and robustness with the aim to develop a standardized neurocognitive profiling battery for animal models of psychiatric disease symptoms. With this focus in mind, in this paper we will not discuss tests that require a lowering of the motivational state by starvation (either food or water deprivation) or extended training periods (such as touchscreen setups for rodents).

Eliciting comprehensive neurocognitive profiles with standardized phenotyping pipelines

The concept of organizing behavioral tests in an arrayed phenotyping pipeline has been realized before, in both rodents and NHPs.[17,18] This approach has several advan-

tages, including requiring fewer animals. It also achieves a high degree of standardization, which in turn improves the reproducibility and comparability of the test battery and allows a dedicated behavioral "subdomain" profile to be generated (see next section). Standardization between institutes allows site effect to be estimated. Moreover, the shortcomings of individual behavioral tests can be ameliorated by having a sufficient degree of redundancy between behavioral measures, which also increases overall robustness. The most important step in designing such a translational neurocognitive test battery is selecting the test paradigms. The main selection criteria to consider are as follows: translatability between species; brevity, so the overall battery is not too long; high test-retest reliability to increase comparability between sites; good balance between effort and predictive validity; and feasibility at different basic research and clinical centers. The time needed for training staff and performing the test, eg, for a learning paradigm, is critical for throughput. Most psychoaffective disorders first appear in adolescence to young adulthood,[19-21] so the maturity of the test animals should match this period in humans.

Cognitive and behavioral subdomain structures

To associate the behavioral measures of such a battery with specific endophenotypes of psychiatric patients, we need to group the tests together to reflect a general neurocognitive profile rather than isolated variables, which cannot be directly translated. Moreover, classifying the tests in this way yields the possibility of reducing dimensions, thereby increasing robustness and attenuating problems of multiple testing.[22]

Several different concepts are used to categorize behavioral measures in translational research in psychiatry. The most straightforward one is to base categorization on the clinical symptoms used in routine clinical diagnostics, such as positive, negative, and cognitive symptoms. This framework has a high face validity but a weak neurobiological basis. Consequently, physiology-focused approaches, such as the Research Domain Criteria (RDoC), were developed.[23] The RDoC define multilevel neurobiological substrates, from biochemical interactions to complex behavior, and consider the stimulation of circuit-based, pro-cognitive mechanisms. The highest-level domains separate behavior into positive valence, negative valence, cognition, vigilance/arousal, and sociability.[23] This system has high content validity but no

direct translatability to cognitive disturbances in psychiatric patients. For the purpose of this review, the definitions of the behavioral domains and classification of the corresponding measures have been modified to balance between construct, content, and face validity, in accordance with previously published concepts.[24-27] The overall behavioral domain structure is separated into positive, negative, cognitive, vigilance/arousal, and social behavior. The RDoC framework is used as a primary reference, but the positive and negative domains refer to clinical symptom categories rather than valences. We use this system to address different cognitive and behavioral phenotypes implicated in neuropsychiatric illnesses; these phenotypes include working memory, social attention, attentional oscillations in perception and performance, sustained attention, response inhibition, proactive and reactive cognitive control, and goal selection.

Assessment of cognition in rodents, nonhuman primates, and humans

To emphasize the translational focus of this review, below we compare and describe the rodent, NHP, and human tests for each of the behavioral domains described in the previous section (ie, positive, negative, cognitive, vigilance/arousal, and social behavior). An overview of the tests can be found in *Figure 1*. Certain tests in rodents and NHPs are used in specific species only; in these cases, the species is named in parentheses.

Positive domain
The most prominent symptoms of the positive domain are hallucinations (ie, visual and auditory perceptions that are not real) and delusions (ie, misinterpreted sensory inputs paired with improper executive functions); these symptoms build key features of SZ and also frequently appear in manic episodes of BD, but they usually do not play a prominent role in MDD. Many different treatment options are available for symptoms in the positive domain. Most of the drugs act on the dopaminergic system, which has been proven as a key modulator of the symptom spectrum (in the form of a hyperdopaminergic state). Nevertheless, it remains virtually impossible to model hallucinations and delusions in animal models. Current models consider increased physical activity, eg, in the open field and Y-maze tests, and alterations of sensorimotor gating—an accepted endophenotype of psychoses—as surrogate tests because they assess changes associated with a hyperdopaminergic state.[27]

The prepulse inhibition (PPI) test is technically simple and robust and is the most commonly used test to assess sensorimotor gating in animals and humans. A complex interplay of feedforward inhibition of cortical and subcortical structures and disturbed E/I balance in many circuits is thought to cause an altered processing of the prepulse, which has an impact on the startle response of the test animal or person. The test consists of a loud, "aversive" tone presented in the

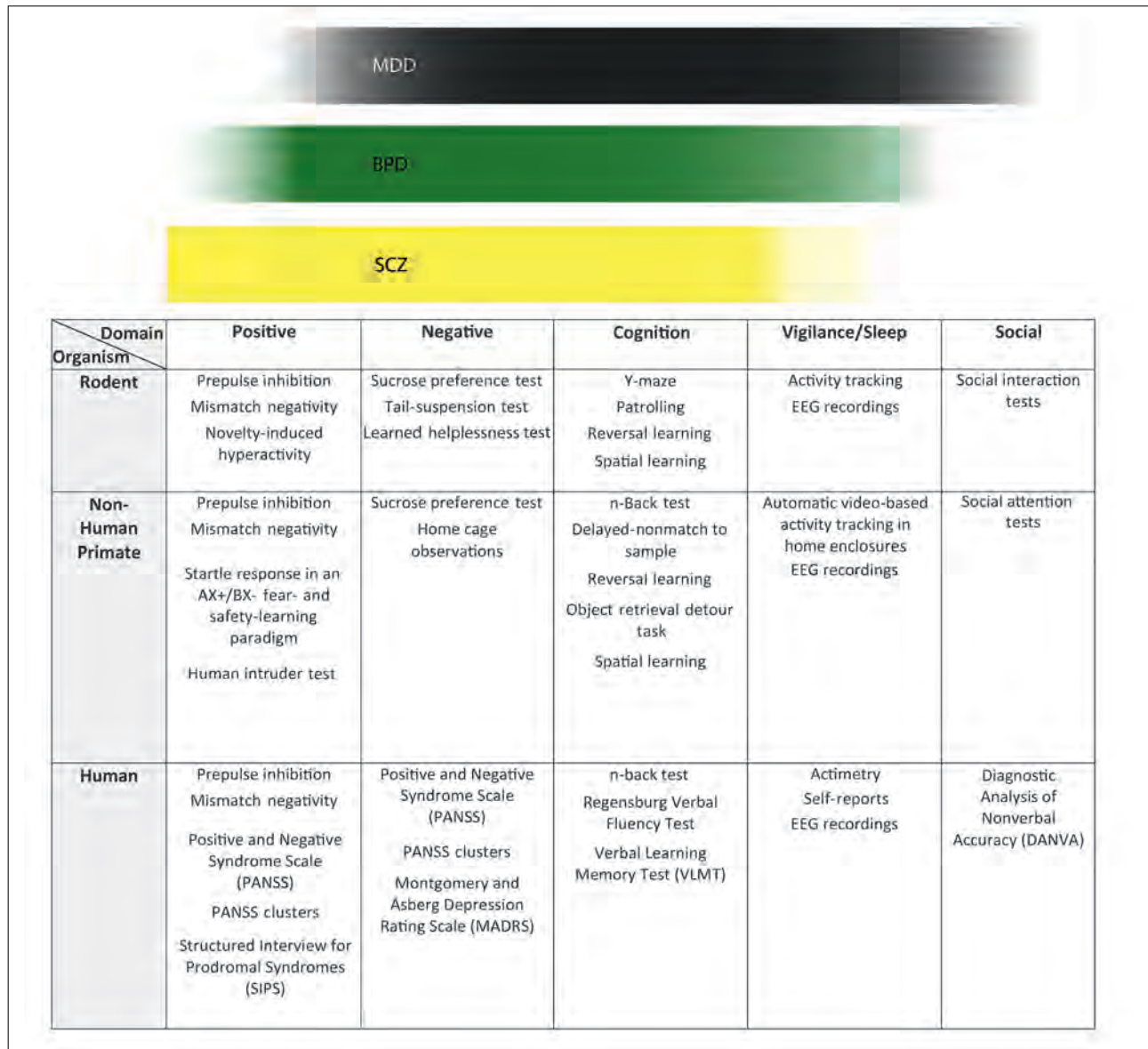| Domain / Organism	Positive	Negative	Cognition	Vigilance/Sleep	Social
Rodent	Prepulse inhibition Mismatch negativity Novelty-induced hyperactivity	Sucrose preference test Tail-suspension test Learned helplessness test	Y-maze Patrolling Reversal learning Spatial learning	Activity tracking EEG recordings	Social interaction tests
Non-Human Primate	Prepulse inhibition Mismatch negativity Startle response in an AX+/BX- fear- and safety-learning paradigm Human intruder test	Sucrose preference test Home cage observations	n-Back test Delayed-nonmatch to sample Reversal learning Object retrieval detour task Spatial learning	Automatic video-based activity tracking in home enclosures EEG recordings	Social attention tests
Human	Prepulse inhibition Mismatch negativity Positive and Negative Syndrome Scale (PANSS) PANSS clusters Structured Interview for Prodromal Syndromes (SIPS)	Positive and Negative Syndrome Scale (PANSS) PANSS clusters Montgomery and Asberg Depression Rating Scale (MADRS)	n-back test Regensburg Verbal Fluency Test Verbal Learning Memory Test (VLMT)	Actimetry Self-reports EEG recordings	Diagnostic Analysis of Nonverbal Accuracy (DANVA)

Figure 1. Translational behavioral and cognitive tests for major neurodevelopmental mental disorders. The figure lists the names of tests that can be applied in rodents, nonhuman primates, and humans to assess behaviors in the social, cognitive, and vigilance/sleep subdomains and in the negative and positive clinical symptom spectrum, as described in the main text. The need for an arrayed set of tests covering these behavioral subdomains is indicated by the shaded and colored bar graphs at the top of the figure, which shows the graded association with major depressive disorder (MDD), bipolar disorder (BPD), and schizophrenia (SZ). For detailed descriptions of the behavioral tests and references, see main text.

presence or absence of a prepulse at a lower intensity; the animal's or test person's subsequent startle response is then monitored.[10] Mismatch negativity is another valuable tool to assess positive symptoms in animals and humans.[28] In this test, a uniform sequence of standard tones is presented that is interrupted by a few (eg, 5%) deviant tones. Responses are measured with an EEG that reflects the amplitude and shape of all the induced auditory potentials. Because the test can be conducted in animals and humans, it has high face and construct validity.[28] In addition to cognitive gating processes, altered exploratory behavior can be an indicator of an overactive dopaminergic system (such a state can also be achieved by administering amphetamine to exogenously stimulate the dopaminergic system).[27] The novelty-induced hyperactivity test is used to assess the natural curiosity behavior of rodents (mice) placed in an unknown environment. In NHPs, the human intruder test can be used to measure anxiety and emotion regulation associated with novelty. Similar to rodents, in NHPs increased reactivity can be associated with an altered stress response modulated by the dopaminergic system.[29] Closely associated with PPI, fear-potentiated startle in NHPs, measured by the startle response in an AX+/BX- fear and safety learning paradigm, has successfully bridged the gap between rodent and human research by modeling emotional regulation, thus increasing the level of translation potential.[25-27,30] In this task, an aversive stimulus inducing a startle response is paired with an auditory or visual cue A followed by cue X (AX+), but no aversive stimulus is presented after a combination of cue B and cue X (BX-).[30]

In humans, several tests use a formalized psychiatric interview to measure positive symptoms. The Positive and Negative Syndrome Scale (PANSS), for example, is routinely used to assess symptom severity in patients with SZ; it consists of 30 items that measure specific symptoms (positive, negative, and general psychopathology), in particular delusions, conceptual disorganization, hallucinatory behavior, excitement, grandiosity, suspiciousness, and hostility on the positive subscale.[31] Furthermore, the Structured Interview for Prodromal Syndromes (SIPS) is a valuable tool to spot the three prodromal syndromes characteristic of people at high risk of developing SZ in the near future.[32] Beyond these interview-based tests, sensorimotor gating can also be assessed in humans by measuring alterations in motor evoked potentials upon paired (or pre-) pulse cortical stimulation, again supporting the high translation potential of the PPI test.[33,34]

Negative domain

Negative symptoms are defined by blunted affect and reduced emotional expression, explicitly in the form of anhedonia (the inability to feel pleasure) and avolition (decreased initiation of goal-directed behavior), both of which are key features of affective and nonaffective psychoses. The clinical spectrum of negative symptoms is highly variable, the underlying neuronal circuits are not well understood, and treatment options are limited.

Various tests are available for measuring negative symptoms. The sucrose preference test (SPT) is an experimentally simple and robust test that is used in rodents and NHPs and does not require specific equipment; it is an appetitive test that measures a sucrose-awarded behavior and is used to assess aspects of anhedonia. The tail suspension test (TST) is a comparably simple, non-appetitive test performed in rodents that measures the animals' intrinsic motivation to escape and thus relates to avolition. Both tests can be applied repeatedly.[25,27] Learned helplessness is another nonappetitive test that assesses coping ability in rodents; however, it cannot be repeated.[35] Different paradigms exist for the learned helplessness test. In general, animals are trained in various setups in which they cannot escape from an aversive stimulus presented in the form of electric shocks. Afterwards, the animals are placed in an environment that gives them the opportunity to escape from these shocks. The extent of avoidance behavior reflects the animals' tendency to learn helplessness in stressful situations, a surrogate for decreased coping strategies. In NHPs (marmosets and rhesus macaques), negative symptom behaviors (eg, social withdrawal) can be assessed by comparing home cage observations with established ethograms.[36,37]

In humans, the negative subscale and the respective negative syndrome clusters of the PANSS can be used to measure negative symptoms.[38-40] The negative subscale and syndrome PANSS clusters comprise the seven most important symptoms, ie, blunted affect, emotional withdrawal, poor rapport, apathetic social withdrawal, difficulty in abstract thinking, lack of spontaneity, and stereotyped thinking. Another tool that can be used in humans is the Montgomery and Åsberg Depression Rating Scale (MADRS), which examines depressive symptomatology. It is a self-administered test that aims at assessing manifestations of depression in nonpsychotic populations.[41] Introduced as a test for depression, the use of the MADRS in the assessment of eg, anhe-

donia and avolition in nonaffective psychoses is justified because of the conceptual overlap between depressive and negative symptoms.

Cognitive domain

Impairments in the cognitive domain as a core feature in affective and nonaffective psychoses often comprise deficits in working memory, attention, executive function, mental flexibility, and declarative/episodic memory. No effective treatment options exist for these deficits, even though they play a crucial role in impaired illness outcomes in affective and nonaffective psychoses.[42] Still, it is promising that various aspects of cognition can be modeled in animals with a level of face validity that enables preclinical treatment trials with a high level of translational potential.[43] Therefore, behavioral profiling should put a special focus on these deficits. In NHPs (marmosets), complex behavioral and especially cognitive phenotypes can be characterized in unrestrained animals in a cage-based cognitive testing system (experimental behavioral instrument, XBI).[44,45] Automatic training in an all-in-one unit mounted to each animal´s enclosure enables experimenters to handle large cohorts even if the tasks are complex. Animals are constantly monitored by several cameras, and the unit also consists of a touchscreen monitor facing the animal´s side, a joystick and response button (similar to modern computer game setups), and a spout for dispensing fluid rewards. Together, these features provide a complex environment for designing multiple behavioral paradigms and measuring the animals' performance. To elucidate the complexity of all the cognitive domains, we will divide them further into several subdomains.

Working memory

The most commonly used test to assess working memory in rodents is the Y-maze test, which uses a Y-shaped chamber; a comparable test, the patrolling test, is performed in the IntelliCage system (TSE Systems GmbH, Bad Homburg, Germany).[46] Both tests are nonappetitive and make use of rodents' natural exploratory and curiosity behavior; mice tend to explore novel areas more frequently in the absence of rewarding stimuli. Movements from "known" to "novel" areas in a maze are monitored for a defined period of time. In NHPs, the n-back test (a modified version of the n-back test for humans) and the delayed nonmatch-to-sample test are established tests of working memory.[47,48] In the latter test, a sample stimulus is presented to the animal; after a short delay, this stimulus is presented together with a novel alter-

native, and the NHP is rewarded for selecting the non-novel alternative. These tests are of special interest within the cognitive domain because of the hippocampus-prefrontal cortex interactions and corresponding deficits, which play a crucial role in the animals' performance. In humans, the n-back test is also used as a simple working memory task. In this test, participants are presented with a sequence of stimuli and asked to indicate when the current stimulus matches the one from several steps earlier. The number of steps after which the matching stimulus is presented can be varied to make the task more or less difficult.

Attention, executive function, and mental flexibility

These three cognitive skills are higher-order cognitive functions that strongly depend on each other; therefore, it is difficult to test them separately. Reversal learning tasks can be used to assess these skills in rodents and NHPs.[49] Altering the reward area as an advanced version of place learning forces the animal to alter its learning strategies towards novel cues, which requires attention and cognitive flexibility; in rodents, these tasks can be performed in water maze and IntelliCage setups. Generally, learning success will be strongly influenced by impaired executive functions; also, increased impulsivity may interfere with test results. Prefrontal cortex function is the most relevant determinant of the animals' performance in these tasks. In NHPs, cognitive flexibility and impulsivity can also be measured by an object retrieval detour task. In this test, a rewarding object is hidden behind a transparent barrier. The prefrontal cortex and its connections to the hippocampus, as well as levels of corticostriatal dopamine, determine success rates in this test.[50] In humans, verbal functioning and thus executive functions can be measured by the Regensburger Verbal Fluency Test, which requires people to access their mental thesaurus under predefined criteria while avoiding repetition and controlling executive processes.[51] In this simple test, participants are asked to produce as many words as possible from either a semantic group (eg, including objects such as food or devices) or a phonemic one (eg, including words with a defined number of syllables). Participants can also be asked to alternate between these two paradigms within the same task, which requires them to shift their attention and thereby tests their mental flexibility.

Declarative or episodic memory

These types of memory are measured in rodents and NHPs with spatial learning tasks followed by probe trials. In the

first phase of the test, the spatial learning task, the animal has to learn a location (eg, associated with a reward); the probe trials then serve to evaluate the animal's memory retention (eg, from which location the reward has been removed). In particular the orbitofrontal cortex, but also frontal areas play an important role in executive functions and the degree of impulsivity and determine learning success in these tasks. Again, in rodents these tests may be performed in the water maze (training) or IntelliCage (preference or avoidance tasks) setup.[52] Only the probe trial can be performed repeatedly. In humans, various aspects of verbal learning and memory can be measured with the Verbal Learning and Memory Test (VLMT), a standardized procedure to evaluate immediate recall, delayed recall, and recognition, amongst other things.[53] After the experimenter has read out a certain number of items from a standard list of unrelated words, participants are asked to recall as many words as possible in any order. The test is usually repeated for (up to) five immediate recall trials and one delayed recall trial. Measures of the participant's performance include the number of items recalled, repetitions, and word intrusions (confabulations). Furthermore, recognition memory capacity can be recorded in separate, simple tasks.

Vigilance/arousal domain

Disturbances of vigilance and sleep are key features of all types of psychoses. Their stability, rhythmicity, and integrity directly influence social functioning and other critical illness outcomes.

In rodents, the IntelliCage records behavior constantly over 24 hours (activity tracking) and therefore can be used to measure circadian parameters and overall activity. Similarly, NHPs (marmosets) can be monitored by automatic video-based tracking in their home enclosures. Additionally, in rodents, NHPs, and humans the qualitative aspects of sleep (REM sleep, delta power, etc) and other circadian aspects can be assessed by EEG recordings through skull-mounted or brain-implanted EEG electrodes in freely behaving animals or head-mounted electrodes in humans. In humans, self-reports of sleep and EEG recordings complemented by transponder-based actimetry are well established paradigms.[54]

Social domain

Social functioning is frequently impaired in affective and nonaffective psychoses. Both reduced and exaggerated social interactions are possible, eg, during depressive and manic episodes in BD.

Social interaction tests in rodents introduce two unfamiliar animals to each other. Mice are very sociable animals by nature and prefer social stimuli to non-social novel objects.[18] Usually, the test mouse is placed into an unknown arena that has already been explored in the adaptation phase by an unknown conspecific, the stimulus mouse. The test can be modified in various ways, but all variations use the extent of the interaction (monitored by video tracking) as a surrogate for social functioning. Various parameters beyond ordinary interaction, such as avoidance, dominance, and aggressive or mating behavior, can be assessed. Experimenters can repeat the test by simply altering the context and stimulus animal. The IntelliCage allows several of the abovementioned social aspects to be continuously monitored.[55] In NHPs, interactions between conspecifics or between NHPs and humans can be observed in social attention tests with different established paradigms, eg, including co-orientation (gaze-following), food sharing tasks, competition tasks, and collaborative tasks. These tasks measure the ability of NHPs to process and act on nonverbal information in a similar way to humans, and they therefore have high translational value. In humans, the Diagnostic Analysis of Nonverbal Accuracy (DANVA) was designed to measure accuracy in sending and receiving nonverbal social information.[56]

The limitations and future directions of translation

The discussion on the predictive value of animal models for psychiatric disorders is ongoing, and many arguments have been advanced for and against the use of such model systems.

One argument is that the high genetic and mechanistic complexity of brain diseases found in GWASs and cellular models cannot yet be appropriately modeled in traditional genetic rodent models. Even modern genetic tools such as CRISPR/Cas9 allow only a few genes to be dysregulated in a single mouse,[57,58] although further progress is expected.[59] Another argument is that psychoaffective disorders are not clearly distinct from each other but represent artificial classifications that have overlapping phenotypes, genetics, and most likely also mechanisms.[6] Therefore, it is not prudent to try to model a specific disease. Each disorder is heteroge-

neous at the level of both clinical phenotypes and genetics, and the situation is likely the same for mechanisms of pathogenesis and recovery.[60] On the one hand, this might at least partially explain the high rates of treatment resistance in patients with affective and nonaffective psychoses, and, on the other hand, any given animal model might represent only a small fraction of patients with a specific disease.

It is highly likely that no disease phenotype exists in rodents that reflects all the major aspects of a psychiatric disorder, and the same might be true for NHPs. There are differences between these model organisms and humans on all levels of biological function. At the cellular level, for example, only one out of ten neocortical GABAergic interneuron (IN) subtypes identified in single-nucleus sequencing in humans corresponds to GABAergic INs in mice.[61] Rodents in particular also show major differences in cortical architecture and functional organization.[62] Cognitive capabilities and complex behavior that are unique to humans, such as speech, must also be considered because they are relevant to cognitive disturbances diagnosed in psychiatric patients but cannot be modeled in animals, ie, neither rodents nor in NHPs. Furthermore, many psychometric tools used to assess symptoms in psychiatric patients rely on self-reports, whereas in animal models cognition can only be assessed indirectly via behavioral phenotypes.

The high attrition rates in drug development for psychiatric disorders after in vivo validation of the drugs in animal models also indicates the low predictive validity of pharmacological effects on cognition and behavior in traditional rodent models.[63] While technological progress is advancing research in human models on the molecular, cellular, and circuit level, we nonetheless still depend on animal models for in vivo studies on cognition and behavior. Thus, we need to improve the construct, content, and face validity of these models and the test paradigms used to assess cognitive function in the hope of enhancing their predictive validity.

The inclusion of NHPs in translational drug validation studies with rodents is a major improvement and is being adopted by more and more laboratories.[64,65] NHPs are a good complement to rodent models because of their high similarity and close evolutionary relationship to humans.

Many neurobiological mechanisms and systems are known to be highly conserved, eg, molecular mechanisms of learning and memory, such as dopamine-dependent neuromodulation and synaptic plasticity in mollusks or serotonin-mediated regulation of social behaviors in crustaceans.[66-68] Hence, it is highly likely that mechanisms complementary to those involved in the pathogenesis of and recovery from psychiatric disorders exist in rodents.

The development of genetic rodent models is also benefiting from new technologies. Tools such as the CRISPR/Cas9-mediated activation and inhibition of the endogenous gene expression of several psychiatric risk genes will simultaneously enhance the construct and face validity of such models for polygenic diseases.[69,70] Although these models will not reflect a substantial portion of the common variants that have been shown to contribute to the risk of developing SZ, for example, a single patient also only carries an as-yet unknown, limited number of these risk variants. Moreover, single nucleotide polymorphisms associated with an increased risk of SZ, BD, and MDD are enriched in regulatory regions of the genome and thus likely converge at the level of pathologically relevant alterations in the expression and/or splicing of an unknown combination of RNAs. If this mechanism is ultimately accepted as the critical primary and causative molecular mechanism of these disorders, enhanced construct validity in genetic models will indeed become a reality in the near future.

Another solution to this issue may emerge as a result of a deeper understanding of the mechanisms underlying the pathogenesis of psychoaffective disorders. Hopefully, research will reveal a point of convergence in the core neurobiological mechanisms that can be modeled at the pathway level in animals. Another approach to recreate the genetic complexity of polygenic brain disorders in rodents are chimeric mice with neural transplants generated from hiPSCs of patients and healthy controls.[71]

These tools are not yet established in NHPs but are being developed and will eventually further complement research on cognition in rodents.[72] Moreover, previous studies successfully used interventions such as manipulation of rearing, application of pharmacologically acting substances, or local lesioning as NHP models of psychoses.[11-13]

Combining genetic rodent models with environmental factors, such as early-life stress (eg, maternal separation) and psychosocial stress (eg, social defeat), not only further

improves the validity of these models, but it also allows for research on gene-environment (GxE) interactions under controlled conditions. These interactions have been shown to play a major role in susceptibility and resilience to psychiatric disorders and in treatment response.[73]

As previously mentioned, these models will probably not mirror the full spectrum of clinical symptoms and mechanistic aspects of specific psychiatric diseases, but they will offer sets of experimentally accessible behavioral subdomains or endophenotypes that correspond to conditions in humans, for example a depressed state in MDD and BD. Such a hypothetical depression-like model would not be limited to the traditional medical classification of psychiatric disorders but might offer a valid model of a class of symptoms found in several psychoaffective disorders.

While some traditional behavioral tests are not easily translated to cognitive disturbances in humans, in recent years more paradigms have been developed with translatability in mind.[74] Experiments such as tests for PPI and mismatch negativity are feasible in rodents, NHPs, and humans and hence have high face and construct validity.[28]

The advent of automated monitoring systems such as the IntelliCage, developed by TSE Systems for experiments with mice and rats, and the XBI system, developed by the German Primate Center for tests with NHPs, allow for testing of a variety of aspects of cognition and behavior under home-cage conditions, which also improves reliability and reproducibility between sites.[45,49]

Another major step forward in the behavioral validation of compounds is the use of several partially redundant tests in a pipeline to ameliorate the limitations of individual paradigms.

Although models and experimental procedures need much improvement to increase the validity and reliability of translational studies on cognition, they remain an essential tool for understanding psychiatric disorders and developing new pharmacotherapeutic compounds to treat them. Such improvements are underway and will hopefully help to push the limits of translation. ∎

References

1. Kessler RC, Angermeyer M, Anthony JC, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry.* 2007;6(3):168-176.
2. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet.* 2018;50(3): 381-389.
3. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018;50(5):668-681.
4. Stahl EA, Breen G, Forstner AJ, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet.* 2019; 51(5):793-803.
5. Bulik-Sullivan B, Finucane HK, Anttila V,

et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015;47(11): 1236-1241.
6. Gandal MJ, Haney JR, Parikshak NN, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science.* 2018;359(6376):693-697.
7. Siskind D, McCartney L, Goldschlager R, Kisely S. Clozapine v. first- and second-generation antipsychotics in treatment-refractory schizophrenia: systematic review and meta-analysis. *Br J Psychiatry.* 2016;209(5):385-392.
8. Hasan A, Falkai P, Wobrock T, et al. World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for Biological Treatment of Schizophrenia, part 1: update 2012 on the acute treatment of schizophrenia and the management of treatment resistance. *World J Biol Psychiatry.* 2012;13(5):318-378.

9. Siskind D, Siskind V, Kisely S. Clozapine response rates among people with treatment-resistant schizophrenia: data from a systematic review and meta-analysis. *Can J Psychiatry.* 2017;62(11): 772-727.
10. Baier PC, Brzózka MM, Shahmoradi A, et al. Mice lacking the circadian modulators SHARP1 and SHARP2 display altered sleep and mixed state endophenotypes of psychiatric disorders. *PLoS ONE.* 2014;9(10):e110310.
11. Gil-da-Costa R, Stoner GR, Fung R, Albright TD. Nonhuman primate model of schizophrenia using a noninvasive EEG method. *Proc Natl Acad Sci U S A.* 2013;110(38):15425-15430.
12. Lavenex P, Lavenex PB, Amaral DG. Spatial relational learning persists following neonatal hippocampal lesions in macaque monkeys. *Nat Neurosci.* 2007;10(2):234-239.
13. Pryce CR, Dettling AC, Spengler M, Schnell

CR, Feldon J. Deprivation of parenting disrupts development of homeostatic and reward systems in marmoset monkey offspring. *Biol Psychiatry.* 2004;56(2):72-79.

14. Brzózka MM, Havemann-Reinecke U, Wichert SP, Falkai P, Rossner MJ. Molecular signatures of psychosocial stress and cognition are modulated by chronic lithium treatment. *Schizophr Bull.* 2016; 42(suppl 1):S22-S33.

15. Gumusoglu SB, Fine RS, Murray SJ, Bittle JL, Stevens HE. The role of IL-6 in neurodevelopment after prenatal stress. *Brain Behav Immun.* 2017;65:274-283.

16. Peña CJ, Kronman HG, Walker DM, et al. Early life stress confers lifelong stress susceptibility in mice via ventral tegmental area OTX2. *Science.* 2017;356(6343):1185-8.

17. Hölter SM, Garrett L, Einicke J, et al. Assessing cognition in mice. *Curr Protoc Mouse Biol.* 2015; 5(4):331-358.

18. Schmitt V, Pankau B, Fischer J. Old world monkeys compare to apes in the primate cognition test battery. *PLoS ONE.* 2012;7(4):e32024.

19. Fink DS, Calabrese JR, Liberzon I, et al. Retrospective age-of-onset and projected lifetime prevalence of psychiatric disorders among U.S. Army National Guard soldiers. *J Affect Disord.* 2016;202:171-177.

20. Williams LJ, Jacka FN, Pasco JA, et al. The prevalence and age of onset of psychiatric disorders in Australian men. *Aust N Z J Psychiatry.* 2016;50(7):678-684.

21. Yin H, Xu G, Tian H, Yang G, Wardenaar KJ, Schoevers RA. The prevalence, age-of-onset and the correlates of DSM-IV psychiatric disorders in the Tianjin Mental Health Survey (TJMHS). *Psychol Med.* 2018;48(3):473-487.

22. Badowska DM, Brzózka MM, Chowdhury A, Malzahn D, Rossner MJ. Data calibration and reduction allows to visualize behavioural profiles of psychosocial influences in mice towards clinical domains. *Eur Arch Psychiatry Clin Neurosci.* 2015;265(6):483-496.

23. Anderzhanova E, Kirmeier T, Wotjak CT. Animal models in psychiatric research: The RDoC system as a new framework for endophenotype-oriented translational neuroscience. *Neurobiol Stress.* 2017;7:47-56.

24. Kaiser T, Zhou Y, Feng G. Animal models for neuropsychiatric disorders: prospects for circuit intervention. *Curr Opin Neurobiol.* 2017;45:59-65.

25. Kas MJ, Fernandes C, Schalkwyk LC, Collier DA. Genetics of behavioural domains across the neuropsychiatric spectrum; of mice and men. *Mol Psychiatry.* 2007;12(4):324-330.

26. Nestler EJ, Hyman SE. Animal models of neuropsychiatric disorders. *Nat Neurosci.* 2010; 13(10):1161-1169.

27. Pratt J, Winchester C, Dawson N, Morris B. Advancing schizophrenia drug discovery: optimizing rodent models to bridge the translational gap. *Nat Rev Drug Discov.* 2012;11(7):560-579.

28. Featherstone RE, Melnychenko O, Siegel SJ. Mismatch negativity in preclinical models of schizophrenia. *Schizophr Res.* 2018;191:35-42.

29. Kalin NH, Shelton SE. Nonhuman primate models to study anxiety, emotion regulation, and psychopathology. *Ann N Y Acad Sci.* 2003; 1008:189-200.

30. Davis M, Antoniadis EA, Amaral DG, Winslow JT. Acoustic startle reflex in rhesus monkeys: a review. *Rev Neurosci.* 2008;19(2-3): 171-185.

31. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull.* 1987;13(2):261-276.

32. Miller TJ, McGlashan TH, Rosen JL, et al. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr Bull.* 2003;29(4):703-715.

33. Berardelli A, Abbruzzese G, Chen R, et al. Consensus paper on short-interval intracortical inhibition and other transcranial magnetic stimulation intracortical paradigms in movement disorders. *Brain Stimul.* 2008;1(3):183-191.

34. Strube W, Bunse T, Nitsche MA, Palm U, Falkai P, Hasan A. Differential response to anodal tDCS and PAS is indicative of impaired focal LTP-like plasticity in schizophrenia. *Behav Brain Res.* 2016;311:46-53.

35. Landgraf D, Long J, Der-Avakian A, Streets M, Welsh DK. Dissociation of learned helplessness and fear conditioning in mice: a mouse model of depression. *PLoS ONE.* 2015;10(4):e0125892.

36. Bauman MD, Lavenex P, Mason WA, Capitanio JP, Amaral DG. The development of social behavior following neonatal amygdala lesions in rhesus monkeys. *J Cogn Neurosci.* 2004; 16(8):1388-1411.

37. Stevenson MF, Poole TB. An ethogram of the common marmoset (Calithrix jacchus jacchus): general behavioural repertoire. *Anim Behav.* 1976;24(2):428-451.

38. Liemburg E, Castelein S, Stewart R, et al. Two subdomains of negative symptoms in psychotic disorders: established and confirmed in two large cohorts. *J Psychiatr Res.* 2013;47(6):718-725.

39. Marder SR, Davis JM, Chouinard G. The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: combined results of the North American trials. *J Clin Psychiatry.* 1997;58(12):538-546.

40. Wallwork RS, Fortgang R, Hashimoto R, Weinberger DR, Dickinson D. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. *Schizophr Res.* 2012;137(1-3):246-250.

41. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979;134:382-389.

42. Galderisi S, Rossi A, Rocca P, et al. The influence of illness-related variables, personal resources and context-related factors on real-life functioning of people with schizophrenia. *World Psychiatry.* 2014;13(3):275-287.

43. Falkai P, Rossner MJ, Schulze TG, et al. Kraepelin revisited: schizophrenia from degeneration to failed regeneration. *Mol Psychiatry.* 2015; 20(6):671-676.

44. Berger M, Calapai A, Stephan V, et al. Standardized automated training of rhesus monkeys for neuroscience research in their housing environment. *J Neurophysiol.* 2018;119(3): 796-807.

45. Calapai A, Berger M, Niessing M, et al. A cage-based training, cognitive testing and enrichment system optimized for rhesus macaques in neuroscience research. *Behav Res Methods.* 2017;49(1):35-45.

46. Bak J, Pyeon HI, Seok JI, Choi YS. Effect of rotation preference on spontaneous alternation behavior on Y maze and introduction of a new analytical method, entropy of spontaneous alternation. *Behav Brain Res.* 2017;320:219-224.

47. Cohen JD, Perlstein WM, Braver TS, et al. Temporal dynamics of brain activation during a working memory task. *Nature.* 1997; 386(6625):604-608.

48. Miller MH, Orbach J. Retention of spatial alternation following frontal lobe resections in stump-tailed macaques. *Neuropsychologia.* 1972; 10(3):291-298.

49. Endo T, Maekawa F, Võikar V, et al. Automated test of behavioral flexibility in mice using a behavioral sequencing task in IntelliCage. *Behav Brain Res.* 2011;221(1):172-181.

50. Jentsch JD, Roth RH, Taylor JR. Object retrieval/detour deficits in monkeys produced by prior subchronic phencyclidine administration: evidence for cognitive impulsivity. *Biol Psychiatry.* 2000;48(5):415-424.

51. Aschenbrenner, S, Tucha, O, Lange, K. W. *Regensburger Wortflüssigkeitstest.* Göttinge, Germany: Hogrefe; 2001.

52. Krackow S, Vannoni E, Codita A, et al. Consistent behavioral phenotype differences between inbred mouse strains in the IntelliCage. *Genes Brain Behav.* 2010;9(7):722-731.

53. Helmstaedter C, Lendt M, Lux S. *Verbaler Lern- und Merkfähigkeitstest.* Göttingen, Germany: Hogrefe; 2001.

54. Wittenbrink N, Ananthasubramaniam B, Münch M, et al. High-accuracy determination of internal circadian time from a single blood sample. *J Clin Invest.* 2018;128(9):3826-3839.

55. Ogi H, Itoh K, Fushiki S. Social behavior is perturbed in mice after exposure to bisphenol A: a novel assessment employing an IntelliCage. *Brain Behav.* 2013;3(3):223-228.

56. Nowicki S, Duke MP, Sisney S, Stricker B, Tyler MA. Reducing the drop-out rates of at-risk high school students: the Effective Learning Program (ELP). *Genet Soc Gen Psychol Monogr.* 2004;130(3):225-239.

57. Wang H, Yang H, Shivalila CS, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell.* 2013;153(4):910-918.

58. Zuo E, Cai YJ, Li K, et al. One-step generation of complete gene knockout mice and monkeys by CRISPR/Cas9-mediated gene editing with multiple sgRNAs. *Cell Res.* 2017;27(7):933-945.

59. Zhou H, Liu J, Zhou C, et al. In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR-dCas9-activator transgenic mice. *Nat Neurosci.* 2018;21(3):440-446.

60. Hodgson K, McGuffin P, Lewis CM. Advancing psychiatric genetics through dissecting heterogeneity. *Hum Mol Genet.* 2017;26(R2): R160-R165.

61. Boldog E, Bakken TE, Hodge RD, et al. Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat Neurosci.* 2018;21(9):1185-1195.

62. Benavides-Piccione R, Ballesteros-Yáñez I, DeFelipe J, Yuste R. Cortical area and species differences in dendritic spine morphology. *J Neurocytol.* 2002;31(3-5):337-346.

63. Pangalos MN, Schechter LE, Hurko O. Drug development for CNS disorders: strategies for balancing risk and reducing attrition. *Nat Rev Drug Discov.* 2007;6(7):521-532.

64. Lameijer M, Binderup T, van Leent MMT, et al. Efficacy and safety assessment of a TRAF6-targeted nanoimmunotherapy in atherosclerotic mice and non-human primates. *Nat Biomed Eng.* 2018;2(5):279-292.

65. McLaren DG, Han S, Murphy BA, et al. DGAT2 Inhibition Alters Aspects of Triglyceride Metabolism in Rodents but Not in Non-human Primates. *Cell Metab.* 2018;27(6):1236-1248.e6.

66. Bacqué-Cazenave J, Cattaert D, Delbecque JP, Fossat P. Alteration of size perception: serotonin has opposite effects on the aggressiveness of crayfish confronting either a smaller or a larger rival. *J Exp Biol.* 2018;221(Pt 12):jeb177840.

67. Kandel ER. The molecular biology of memory storage: a dialogue between genes and synapses. *Science.* 2001;294(5544):1030-1038.

68. Tierney, A. J, Hanzlik, K. N, Hathaway, R. M, Powers, C, Roy, M. Effects of fluoxetine on growth and behavior in the crayfish Orconectes rusticus. *Mar Freshw Behav Physiol.* 2016;49:133-145.

69. Cheng AW, Wang H, Yang H, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* 2013;23(10):1163-1171.

70. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc.* 2013;8(11):2180-2196.

71. Chen C, Kim WY, Jiang P. Humanized neuronal chimeric mouse brain generated by neonatally engrafted human iPSC-derived primitive neural progenitor cells. *JCI Insight.* 2016;1(19):e88632.

72. Qiu Z, Li X. Non-human primate models for brain disorders - towards genetic manipulations via innovative technology. *Neurosci Bull.* 2017;33(2):247-250.

73. Gonda X, Petschner P, Eszlari N, et al. Effects of different stressors are modulated by different neurobiological systems: the role of GABA-A Versus CB1 receptor gene variants in anxiety and depression. *Front Cell Neurosci.* 2019;13:138.

74. Hvoslef-Eide M, Nilsson SR, Saksida LM, Bussey TJ. Cognitive translation using the rodent touchscreen testing approach. *Curr Top Behav Neurosci.* 2016;28:423-447.