Confusing preclinical (predictive) drug screens with animal "models" of psychiatric disorders, or "disorder-like" behaviour, is undermining confidence in behavioural neuroscience

S.C Stanford

Department of Neuroscience, Physiology & Pharmacology,

University College London

Gower Street,

London WC1 6BT

c. stanford@ucl.ac.uk

Abstract:

Preclinical (predictive) screens for psychotropic drugs are often used, incorrectly, as animal "models" of psychiatric disorders, or to study "disorder-like" behaviours. This misunderstanding is contributing to poor translation and is undermining confidence in behavioural neuroscience. In this editorial, I discuss some of the reasons why the interpretation of results from many of these procedures is dubious because the criteria for validity of the test, as a model of the disorder, have been ignored. Arising from this, I propose that the description of any abnormal behaviour of rodents as a "model" of a psychiatric disorder, or even "disorder-like", without evidence-based justification, should be regarded as unacceptable in this journal.

[Editorial text]

A distinguished colleague commented recently that neuroscience papers in high impact journals are now often “decorated with some behavioural data in order to strengthen their translational relevance”. Unfortunately, behavioural neuroscience stands accused of a poor record for translation and the accumulating disappointments are undermining the entire field. Poor experimental design, subjective bias and inappropriate statistical analysis are currently attracting the most attention as causes of the problem. However, in many cases, the explanation could arise from flawed interpretation of the behavioural data. For reasons I explain below, I propose that this journal takes the lead in no longer condoning automatically the description of any abnormal behaviour of rodents as a model of X-disorder, or ‘X-disorder-like’, because such claims often carry little, if any, scientific justification.

Experience has confirmed that procedures such as Porsolt’s Forced Swim Test, the Elevated Plus-Maze, the Light-Dark Box, the Open Field and many others can be efficient screens of compounds when looking for new antidepressant or anxiolytics drugs (of certain chemical classes). The problem is that many researchers go on to assume that, in the drug-free state, these tests can be used as “models“ of depression or anxiety. Equivalent assumptions are made for other psychiatric illnesses: a particular bête noir of mine is hyperactivity in the Open Field, which is now being used as a “model of bipolar disorder”. Such claims show a fundamental misunderstanding of the status of predictive drug screens (see: Willner 1984), which have never been validated as animal models of anything. As a consequence, the scientific literature is being progressively riddled with dodgy conclusions.

A specific example of this violation is when animals that are the first to give up swimming in the Forced Swim Test are regarded as expressing something akin to depression. Another is when it is inferred that animals that spend little time on the open arms of the Elevated Plus-Maze are anxious. This is equivalent to concluding that, if all drugs that make rats whistle the national anthem turn out to be antidepressants in humans, then animals that do not express this behaviour in the drug-free state are depressed. To put it another way: it is often claimed that a deficit in struggling in the Tail Suspension Test reveals depression-like behaviour but, if two humans were suspended upside-down, against their will, I doubt that a difference in their struggling would be attributed to one being more depressed than the other.

Neither of these points has any bearing on the validity of these tests as screens for antidepressant drugs. It is even possible that by increasing motor activity, which is an action shared by all antidepressants in these tests, the response to these drugs translates directly to humans as a beneficial effect on motor retardation in depression. Whether or not this is the case, it cannot be inferred that stress-induced immobility in these tests is a rodent version of depression.

To be pragmatic about this, there are more than twenty diagnostic features of depression in humans but neither a motor deficit, when confined within a tank of water, nor a reduction in struggling, when suspended upside-down, is on the list. Similarly, anti-anxiety drugs increase activity on the open arms of the Elevated Plus-Maze but most people would regard any reluctance to 'walk the plank' (i.e., to venture onto the open arms of the maze) in the drug-free state as a sensible decision after a risk assessment, rather than a sign of a psychiatric illness. Indeed, increased activity on the open arms could indicate reckless or impulsive behaviour. As for the Open Field (and other tests that rely on contrasting environments), it is rarely possible to tell what is being evaluated because the interaction between emotionality and ambulation, and the possibility that the experimental intervention has altered the animals' sensitivity to light, are usually ignored altogether (Stanford 2007). We should also be wary of the recent rebranding of some behavioural screens. For instance, I note that the Social Interaction and Marble-Burying tests, which have been used for decades to screen anxiolytic drugs, are now being used to score poor social function and perseveration of mutant mice in preclinical research of autism.

Many authors agree that describing an abnormal behaviour in rodent preclinical tests as a "model" of a psychiatric disorder is stretching anthropomorphism too far and flag this insight by describing them as 'disorder-like' instead (i.e., depression-like, anxiety-like….). This is disingenuous: it implies that authors are fully aware that scores for these behaviours depend on the animals' environment (unlike psychiatric disorders, with the exception of specific phobias) and disappear when the animals go home (unlike psychiatric disorders, with the possible exception of agoraphobia). However, by describing the behaviour of interest as "X-disorder-like", researchers are hedging their bets and trying surreptitiously to persuade readers that the behaviour they have evaluated is analogous to the psychiatric illness they are researching.

As implied above, the possibility that environmental context influences animals' behaviour should always be regarded as a relevant factor, not least because most procedures impose some degree stress

on the animals. In some cases, this involves merely exposure to a novel environment and separation from its cage-mates. For others (e.g., the Morris Water-Maze), the validity of the test relies on the stress being severe enough to motivate the animals to perform the task as quickly as possible. What we need to know is how animals behave when they are not exposed to such threats. An ethological study in their natural environment would be ideal, but studies of animals that are group-housed in their home cage would be a step in the right direction. So far, there is no way of profiling multiple aspects of the behaviour of an individual under these conditions, still less interactions between cage-mates, but promising technologies are in an advanced stage of development.

Another limitation, mentioned above, is that psychiatric disorders are diagnosed on the basis of a constellation of problems (of different weightings), not a single feature. Moreover, different combinations of symptoms and signs in humans can all qualify for the same clinical diagnosis. This means that even conscientious scientists, who carry out a routine battery of three or four different behavioural tests (which usually merely evaluate different types of motor response, but not mood), cannot claim to be satisfying any diagnostic checklist. For people who use only one test, there is the additional confounder that this might, at best, reveal an endophenotype, which could be evident in several different psychiatric and neurological disorders. For instance, a deficit in social interaction / social withdrawal is evident in depression, social phobia, schizophrenia and autism, while perseverative behaviour is seen in anxiety (OCD), attention deficit hyperactivity disorder and autism.

Obviously, the definitive way to validate an animal model of a psychiatric disorder would be through back-translation of a specific (human) biomarker. Unfortunately, biomarkers are on psychiatrists' wish-list too. Until then, the validation process rests on finding combinations of multiple features of the model that best resemble those seen in humans suffering from the disorder of interest. These would include behavioural, neurochemical, hormonal, pharmacological and, increasingly, genetic factors. A good example of that approach is the olfactory bulbectomized rat, as a model of depression (Kelly et al., 1997), even though we should bear in mind that that depression is not explained by olfactory bulbectomy in humans.

Probably the biggest gremlin is the blinkered assumption that the experimental intervention is the direct cause of the behavioral change of interest. For instance, if a gene knockout exacerbates immobility in the Forced Swim Test, it is usually concluded that the gene is responsible for preventing the immobility.

The possibility is rarely considered that the exact opposite is the case: i.e., that expression of the gene could be responsible for promoting movement when animals are stressed, rather than preventing stress-induced immobility, and that these are two neurobiologically distinct processes.

Furthermore, by analogy with the complementary neuronal networks that regulate appetite *versus* satiety, sleep *versus* arousal or pain *versus* analgesia, it is highly likely that networks that promote depression *versus* antidepression, or anxiety *versus* antianxiety, have similarly dedicated functionality. This point reflects the proposal that therapeutic drugs do not target the cause(s) of the psychiatric disorder, but bypass the locus of the problem and restore normal function by following a different route through the brain (see: Reid and Stewart 2001). This could explain the different features of relief of depression by ketamine and classical antidepressant drugs, for instance: whereas one switches off depression, the other switches on antidepression. This dissociation is why a monoamine theory of antidepression is still viable, despite the monoamine theory of depression now being regarded as implausible or, at best, a special case.

These disparate possibilities need to be considered as a matter of routine: i.e., that a behavioural change following an experimental intervention (e.g., gene knockout) could arise from either promotion of stimulation of one response, or loss of inhibition of another (or *vice versa*) and that it is possible that neither of these interpretations has any bearing on the cause of the disorder or its relief following drug treatment. In short, to restore confidence in the conclusions from preclinical behavioural procedures, we must ensure that the results are interpreted cautiously and that all possible explanations are considered - not just the one that appears to fit the preferred hypothesis.

All these points will be familiar to preclinical psychopharmacologists who, as readers of this journal, will certainly think that this article is preaching to the converted. However, we evidently need to be far more rigorous in ensuring that misleading and partial science does not slip through the peer review net. To thwart the skeptics (who have a point) from succeeding in throwing out our baby with their bath-water, I propose that this journal should not publish any paper that describes an abnormal behaviour as a "model" or 'X-disorder-like' unless there is a consensus among the editors that such a claim has been plausibly validated. I am confident that few, if any, will qualify. For those that do not, the behaviour should be described in terms of exactly what it is: e.g., swimming (stress)-induced immobility; open-arm activity; ambulation in the light arena, stress-induced anhedonia…. An example of how this reductionist

approach could make a crucial difference to whether or not a drug action in rodents translates to the target therapeutic application in humans is explained in Stanford (2014). By being more circumspect in the interpretation and validation of behavioural 'models', and peer review of such studies, we might ensure that promising findings that are reported in the *Journal of Psychopharmacology* have a better record for successful translation than those in some other journals with stellar impact factors.

References

Kelly JP, Wrynn AS, Leonard BE (1997) The olfactory bulbectomized rat as a model of depression: an update. *Pharmacol Ther* 74:299-316.

Reid IC and Stewart CA (2001) How antidepressants work: new perspectives on the pathophysiology of depressive disorder. *Br J Psychiatry* 178:299-303.

Stanford SC (2007) The Open Field Test: reinventing the wheel. *J Psychopharmacol.*21:134-5.

Stanford SC (2014) Psychostimulants, antidepressants and neurokinin-1 receptor antagonists ('motor disinhibitors') have overlapping, but distinct, effects on monoamine transmission: the involvement of L-type Ca2+ channels and implications for the treatment of ADHD. *Neuropharmacology* 87:9-18.

Willner P (1984) The validity of animal models of depression. *Psychopharmacology* (Berl) 83:1-16.