Back to the future II

Validation of paradigms of the past and technology of the future.

Prof. B.M. Spruijt

Delta Phenomics, Utrecht, the Netherlands and
Department of Biology, Utrecht University, Utrecht, the Netherlands.

The first open field study for rodents was conducted 1934; the issue of what is exactly measured in such an open field test has been addressed in the first review written in 1973 [1]. In a recent review Haller & Alicki [2] mention that the open field is still a frequently used test nowadays, despite innovative alternatives. The open field is used for exploratory behaviour, habituation, anxiety depression, schizophrenia, but also for more pharmacological questions. The ethogram used has always been a topic of debate. Technological innovations in molecular biology, biochemistry electrophysiology and histochemistry have enormously facilitated the level of resolution of independent variables, the acceleration of production of results and yielding new read out parameters and new insights. Now, the question emerges: Are there comparable benefits from technological innovations for behavioural science in terms of new parameters, results and insights as well? What has changed in the course of 80 years?

The introduction of the Skinner-box with its increasingly automated electro-mechanical devices is indistinguishable connected with the behaviouristic approach and has induced a discipline on its own already a long time ago. According to a review of Haller and Alicki [2] all innovations are mainly used by the innovators themselves. In fact, the oldest tests are still most frequently used, almost in the same way for decades. Has it just been a mere improvement in efficiency and costs replacing the human hand and eye by a device? This is more than a rhetoric question.

The topic of this paper highlights the necessity to validate novel technologies in behavioural sciences, an attempt that is hampered by the fact that the classical behavioural tests are poorly validated.

The driving force behind the use of well-known tests for studying the proximate causation of behaviour is the scientific discipline of comparative or translational research which uses animal models for human health and medical care. Of all 2.637.720 behavioural studies (including economics etc.) 775.662 are on human behaviour, 123.726 are on rat behaviour and 59.320 are on mouse behaviour. So by far, animal behaviour means 'rodent' behaviour studied as model for man in neuroscience and preclinical research.

Oversimplified ethological set ups are characteristic for laboratory tests: next to Skinner boxes, various mazes, open field, light dark boxes, conditioned fear shock box, defensive behaviour paradigms, attention tasks, all are classical tests. While the Skinner box is limited in its underlying assumptions resulting in conceptual constraints, ignorance of genetically determined natural behaviours and focus on a few artificial read out parameters (nose pokes, lever presses etc.), most other tests suffer from biased human observations and short test durations. Often, the focus is on behaviour of individuals, mostly of the male sex. We may assume that humans may be skilled in interpreting primate behaviour, but the nocturnal rodent who is strongly relying on olfactory senses, touch and also sound – this is a different story. Furthermore, humans can hardly resist interpreting behaviour while observing. For example, when the distance between two animals declines, the human observer easily scores 'approach behaviour" as if he/she can recognize the intention of the animal.

The subjective influence of the human observer and the large inter-observer and inter-lab differences are substantial. The ongoing debate on the fuzzy and hardly reproducible results culminated in papers in the nineties and early 2000 when mouse mutants became available in large numbers [3].

Automated methods could solve the reliability issue. But then scientists rightfully demand validation of these methods and techniques as there is no a priori 'intuitive' knowledge of a specific behaviour defined by a computer algorithm.

To understand why some tests have been successfully automated and others not one has to distinguish two types of research question using these tests:

- A test for addressing a straightforward question where a few behavioural read-out parameters are used to assess dose-response relationship. Example: amphetamine induces locomotor activity. Distance moved is then used as a litmus paper indicative for the effective dose. This approach is applied in behavioural and safety pharmacology.

- The same parameters in the simple test are used for the functional interpretation of a treatment. Let us say the same test is used for the efficacy of diazepam in anxiety. Now low activity (immobility or freezing) is assumed to be indicative of anxiety and the drug acts anxiolytic by increasing locomotor activity.

- Tests used to answer straightforward questions are not necessarily validated for functional interpretations. Thus, the validity of a specific set up depends on the scientific question and the interpretation of behaviour depends on the scientific context. Nonetheless, we see whole batteries of tests (mainly simple tests previously used for rats) being used for describing the effects of a novel drug or the phenotype of unknown mouse mutants in functional terms.

Given the broad array and the huge number of studies conducted over time using those classical tests of open field and elevated plus maze, the need for validation seems to have faded away against the background of a familiar and extensive mixed set of data. The interpretation of immobility in case of an anxiolytic drug action and low activity in case of a stimulant may be acceptable when the treatment effects are well known.

The question remains: How can behaviour such as immobility be used for a biological relevant functional interpretation of behaviour in terms of anxiety when an unknown drug or mutant mouse is studied?

The validation of innovative technology is difficult to design in a convincing way because of unclear validation criteria. Let us consider two aspects:

- First, the aim is to get rid of human (mainly visual) observations. However, there is that desired link with the previous intuitive way of scoring and interpretation. We want a new technology to score immobility the same way as human observers do, even when inter-observer reliability is low. We have to accept that if new technology has to match previous methods, little will be gained. For straight forward questions such as: amphetamine induces locomotion, previous results will be confirmed.

- Second, novel technologies allow extending the time of observation and ethograms. Novel drugs or mutations may have all kind of effects on behaviours which have not been measured before, complicating the previously assumed and simplified specificity.

The caveat of a more automated comprehensive approach, detecting a wider spectrum of behaviours over a longer period of time is that it yields a complex set of data. Those numbers representing movements, velocity, (changes in) contour of animals etc. do not have a direct intuitive representation in terms of behaviours as one is familiar with by using classical studies. Thus, at first sight the data set might be apparently without any meaning.

This may be business as usual in other scientific disciplines where math is more generally used and accepted, but cumbersome for behavioural scientists, who like to rely on their own way of observing animal behaviour which leads us back to the first item.

How to get out of this vicious circle?

It requires special tools for exploring the data. Descriptive studies require special data mining tools. Hypothesis driven studies have *a priori* defined parameters and those which appeared to have changed without *a priori* prediction need to be explained. Since the number of parameters may be very large, an appropriate analysis and interpretation is intriguing and challenging. Behaviour as human observers see it, always elicits associations in terms of meaning or differences. Scepticism and doubt emerge about the meaning of those newly generated abstract numbers. One is tempted to fall back on familiar and 'simple' tasks.

The user – not the observer in case of automated observations – has to rely on numbers, not knowing exactly how these numbers are generated. Therefore, the need for proper validation is justified. But this requires an appropriate test paradigm which allows the detection of the parameters generated by the technology.

It is the tradition and the apparent need for simplicity which have restricted the validation and use of automation and technological innovation.

Now, automation of the measurement of parameters in classical tests focusing on individual animals in a relatively simple environment lent itself for automation and it is there where technology found its entrance.

The advantage is mainly a matter of saving time, more precision etc. but hardly any new parameters. Distance moved and time spent in certain areas are probably the most commonly used parameters.

For the progress of behavioural sciences in methods and paradigms we need the two available kinds of technical innovations:

- Statistical and methodological procedures to go beyond the mere counting of frequencies and durations.
- Data collection instruments by use of sensors connected to a computer.

ad1) That behaviour is more than a sum of frequencies and/or duration had been noticed before: behaviour is a stream of events organized in time and space: temporal organization or sequential organization of behaviour as is explained in various ethological text books. Both time and space had been neglected in the most frequently used tests. The elegant studies of Golani and Benjamini [4] show that exploration is a gradually extending patterning of behaviours in the available space, which has to be bigger than anything regularly used in laboratories. The analysis of the temporal organization by using transition matrices or t-patterns had also been used decades ago, but did not evolve into a general tool. The rapid access to computer power eased the use of these methods but even that did not yield a breakthrough of such methods. However, mainstream users sticking to classical methods slow down the progress in this field.

ad 2) Infra-red beam devices are a simple way of monitoring activity; a kind of grid and the interruption of a beam could allocate the animal to certain area in an open field-like apparatus. Later on, video- images were digitized and then fed into a computer. These methods are activity based thus, quantifying a change in position of an animal. Another set of features was added later. The animal was now not reduced to a single point such as for instance the centre of gravity, but the nose tail points of the animal and the contour of its periphery was used to include more form related characteristics. The dynamic changes of the contour allowed the distinction of some behaviours such as grooming, rearing exploration, immobility etc. The need for high throughput phenotyping tried to take advantage of automated devices. However, well known classical tests were often just combined in a battery of tests and therefore counteracting the initial approach.

The full potential of using technology which could bring the whole discipline of behavioural science to a higher level requires a paradigm shift as the animals have to be provoked to display the whole potential of behaviours. Though, technology now allows a shift in paradigms as the laborious and complex human observations can be avoided and new parameters can be added, it requires validation of the paradigm and the technology.

This is one of the reasons for the lack of progress in using new paradigms in behavioural neuroscience.

Pharmacological validation asks for the use of reference drugs. Thus, if a new system with the capacity to measure anxiety related behaviour is tested, the antagonistic effect of e.g., diazepam has to be demonstrated first. In contrast, effects of diazepam are based on the limited technology one wants to improve. A new technology might show that anxiety is more than immobility and that diazepam has more effects than counteracting immobility. It is not easy to break through this circular reasoning, unless it's accepted that diazepam has more effects than previously assumed and that it is not the ultimate reference drug. For testing anxiety, consensus on what anxiety is under certain conditions is required. Therefore, we may have to return to and upgrade ethological concepts.

Advanced analyses have been developed and novel measuring techniques have been introduced. There is a huge opportunity for improvement in the main stream of behavioural sciences. The strength of automated methods is that one has to explicitly formulate the behaviours of interest before they can be translated into a computer algorithm.

A machine has the ability to perform with sustained attention and, thus, the factor time can now really be involved in the stream of behaviours. Rhythmicity can be revealed and behaviour of nocturnal animals can easily be monitored during the dark phase of the circadian rhythm. Surprisingly and to our disappointment, extending the duration of testing is not really extensively used yet.

Another issue is the reproducibility. Whereas inter-observer reliability is low, a machine scores the same stream of events almost identically to the advantage of reliability.

We have to demonstrate that new methods of observation and analysis tools catalyse the use of new paradigms yielding more and biologically relevant information, stimulating cross fertilization between innovative technology and expertise in behavioural sciences, an acceleration of its development is the near future. Studies will be conducted under biologically-relevant conditions. This will yield animal models with enhanced translational value in preclinical research, new main effects and unwanted side effects of drugs can be detected within the same paradigm. If we accept the limitations of existing paradigms and accept that behaviour is more complicated than previously assumed then technology may help to explore and understand behaviour in depth and stimulate for instance drug discovery to contribute to human health.

References

1. Kulkarni SK, Dandiya PC. (1973). Effects of antidepressant agents on open field behaviour in rats. *Psychopharmacologia* 33, 333-338
2. Haller J, Alicki M (2012). Current animal models of anxiety, anxiety disorders, and anxiolytic drugs. *Curr. Opin. Psychiatry.* 25, 59-64.
3. Wahlsten D, Bachmanov A, Finn DA, Crabbe JC. (2006) Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc. Natl. Acad. Sci. U. S. A.* 103, 16364-16369.
4. Golani I, Benjamini Y, Eilam D. (1993). Stopping behavior: constraints on exploration in rats (*Rattus norvegicus*). *Behavioural Brain Research* 53(1-2)21-33.