

## Behaviour and the standardization fallacy

Behavioural phenotyping of mouse mutants has become an important tool in the search for biological roles of single gene products in complex mammalian traits, such as memory, aggression, anxiety and addiction (see page 277 for more on the genetics of addiction<sup>1</sup>). Because specific strains of mice, husbandry practice and test protocol may influence behavioural effects of mutations in unpredictable ways, scientists strive for consensus on rigorous standards to maximize reproducibility of results across laboratories<sup>2</sup>. Increasing reproducibility of results through standardization, however, accentuates and obscures the very problem—that of reporting artefacts that are idiosyncratic to particular circumstances.

Standardization serves to reduce individual differences within study populations (within-experiment variation) in order to facilitate detection of treatment effects, and to reduce differences between studies (between-experiment variation) in order to maximize reproducibility of results<sup>3</sup>. If standardization were fully effective, inter-individual variation within study populations would decrease to zero, and reproducibility of effects between experiments would increase to 100%. At the same time, however, each experiment would turn into a single-case study with zero information gain. This is what I here refer to as the standardization fallacy: the increase of reproducibility

at the expense of external validity.

External validity stands for “how applicable your results are to other situations (environmental contexts), populations or species”<sup>4</sup>. External validity is an inherent feature of a result and will not be affected by standardization. However, standardization increases the risk of detecting effects with low external validity (or of missing effects with high external validity). In contrast, reproducibility can be increased simply by equating situations more carefully and, hence, tells us nothing about external validity. A result that is highly reproducible under highly standardized conditions may therefore poorly generalize to other conditions, whereas high external validity necessarily goes together with high reproducibility, even when conditions are poorly equated between replicate studies.

Of course, one might argue that it is irrelevant whether a genetic effect reflects a specific genotype by environment interaction or a highly penetrating genetic effect. Both may be biologically real, theoretically meaningful and, hence, informative. This is certainly true if the experimental design allows one to generate ideas about the nature of the interaction as, for example, in a recent study by Tsien and colleagues<sup>5</sup>. They showed that certain memory deficits induced by the lack of NMDA receptors in a subarea of the

hippocampus are overcome by environmental enrichment, possibly through enrichment-induced, NMDA-independent synaptogenesis. In contrast, in another study<sup>6</sup> that compared several mouse strains in three different laboratories under conditions that had been rigorously equated across sites, the lab-specific effects of genotype on many of the behavioural measures had to be attributed to unintended sources of variation, and hence were, by definition, artefacts. It is important to note that using a single standardized genetic or environmental background or test situation for the characterization of mutants makes it impossible to distinguish artefacts from informative effects. Systematic variation of situations is the only means to determine the nature, and demonstrate external validity, of genetic effects on behaviour. As these considerations also apply to treatments, such as drugs and brain lesions, systematic variation of situations should form an integral part of all animal experimentation.

Hanno Würbel

*Institute of Animal Sciences, ETH Zürich,  
Schorenstrasse 16 / SLA B14, 8603  
Schwerzenbach, Switzerland (e-mail:  
hanno.wuerbel@inw.agrl.ethz.ch)*

1. Nestler, E. *Nature Genet.* **26**, 177–281 (2000).
2. Wahlsten, D. in *Symposium: Behavioural Phenotyping of Mouse Mutants* (ed. Office of the Associate Dean of Scientific Affairs, University of Cologne, Germany) 15–16 (2000).
3. Beynen, A.C., Gärtner, K. & van Zutphen, L.F.M. in: *Principles of Laboratory Animal Science* (eds van Zutphen, L.F.M., Baumans, V. & Beynen, A.C.) 101–108 (Elsevier, Amsterdam, 1993).
4. Lehner, P.N. *Handbook of Ethological Methods*, 2nd edn (Cambridge University Press, Cambridge, 1996).
5. Rampon, C. *et al. Nature Neurosci.* **3**, 238–244 (2000).
6. Crabbe, J.C., Wahlsten, D. & Dudek, B.C. *Science* **284**, 1670–1672 (1999).