A critical reflection on temperament assessment tests in dogs

Author: Helen Brown

Temperament assessment in companion animals is an important welfare issue, and more so for dogs than any other species. In this essay the purpose and process of the tests will be outlined, followed by a discussion of the potential positive and negative welfare implications. This will focus largely on the current weaknesses in the application of these assessments and the ethical and welfare concerns implicit in this. Suggestions will be made for how the procedure and its outcome could be improved.

Temperament assessment: what, how, why, when and who?

Temperament is a description of stable behavioural tendencies, and is a function of genetics and acquired experiences. In the domestic dog, temperament tests are intended to assess behaviour and predict long term traits so as to establish suitability for a particular purpose. The tests may take the form of ethograms, graded responses to different stimuli (e.g. toys, people, other animals), responses to handling or even questionnaires (Wiener and Haskell, 2016). They can be done at any age, but have traditionally been applied to younger dogs in order to make a decision on their future training and working roles- for example, assistance, herding or retrieving dogs. Anyone can attempt to perform a temperament test, but it is a skilled task requiring knowledge and training in canine ethology (Rayment *et al*, 2015).

Strengths and weaknesses

The biggest potential strength of temperament testing from a welfare point of view is the ability to apply it to adult dogs in re-homing shelters. Behaviour is a crucial factor in being re-homed successfully, being kennelled for a long period or even being destroyed. Reducing the post-adoption relinquishment rate and euthanasia rate is a positive outcome. However, when life-ending decisions have to be made, it is good practice for them to be strongly evidence-based, and this is where temperament testing is weak.

Currently there is no agreed method of testing, a lack of consistency in application, and not enough evidence on test validity (Taylor and Mills, 2006). Even the terminology around these tests is inconsistent (Diederich and Giffroy, 2006), and useful information from the field of human psychology has not been applied significantly (Rayment *et al,* 2015). The influence of the context in which testing takes place is a particular concern for shelter dogs, both because of the environment where testing takes place and its impact on results, and the consequences for the dogs being tested. A lack of standardisation is a significant welfare problem for dogs (Mornement *et al,* 2014) and the process of testing can be a potential source of adverse stress in itself.

Temperament tests are rife with questions over reliability and validity. In guide dogs, testing is part of the breeding programme where genetics and environment are relatively consistent, and high predictability has been found (Harvey *et al*, 2016 and Duffy and Serpell, 2012). In other working dogs the findings vary widely- from Border Collies with low predictability (Riemer *et al,* 2014) and military dogs where findings were variable (Arvelius *et al,* 2014 and Sinn *et al,* 2010).

The difficulty in making meaningful comparisons is that these working dogs were all being assessed in different ways for different jobs. Also, there is a difference between predicting suitability for a job than from predicting stable future behavioural traits (Willson and Sinn, 2012). In animals in general, behaviour is generally highly predictable (Bell *et al,* 2009), but not only do domestic dogs vary so much, so does the information generated about them by a wide variation in test protocols. Tests that are predictive of one thing in one type of dog cannot be assumed to extrapolate to dogs of differing breeds and ages. When the tests applied are varied, prone to observer bias, inaccurately measured, and performed in a stressful environment, it is hard to justify the life-changing (or life-ending) consequences for an animal.

Better applications

The first meta-analysis of personality consistency in dogs found that there is significant consistency and that it is higher in older dogs (Fratkin *et al,* 2013), potentially advantageous for the prospects of shelter dogs. If this is true, then the first duty of those who make decisions using temperament tests is to establish that the behaviours measured are in fact representative of 'personality'. This requires expert data collection, consistency of testing, publication and peer review, with ongoing refinement. As it stands in 2017, temperament tests are often not based on established canine behaviour science. Other researchers (Dowling *et al*, 2011 and Valsecchi *et al,* 2011) have found good predictability in shelter dogs, but again the trait dimensions assessed and test methods differ from one data set to another. A diversity of backgrounds of research into this field contributes to difficulty of reaching an agreed standard (Jones and Gosling, 2005).

Ethical issues

In human psychology, 'personality' is a contested concept (Butt, 2004), and the popularity and dominance of psychometric testing in human culture could arguably have contributed to high expectations being placed on the value of temperament tests in domestic dogs. This might apply both to the professionals doing the testing, and a potential owner who believes their dog will always behave 'well' or at least consistently, never developing a behavioural disorder. Even the most sophisticated test protocol will be imperfect, and it is an ethical concern if dogs are seen to 'fail' a temperament assessment when a moral overtone is attached to canine behaviour.

The question also has to be raised of whether human society should be making these management choices about dogs. Pragmatically, a good application of the tests is likely to lead to welfare improvement, both for

shelter dogs and, in a better regulated industry, for dogs bred as companions. But is breeding or selecting "low maintenance" dogs in temperament terms analogous to genetically engineered "pain-resistant" meat animals? Perhaps this might contribute to a negative overall perception of the purpose of animals in general and an excuse not to try harder to understand canine behaviour in particular.

Conclusion

It is clear that temperament assessment tests in dogs are highly flawed, with inconsistent testing regimes, poor understanding by some testers, and grave concerns over reliability and validity. The welfare consequences can carry high stakes for the dogs in question. Whether humans should be making decisions for animals at this level is a moot point. In a world where animals are still widely considered as tools for human use, the most positive welfare outcome with respect to domestic dogs and temperament testing is to work towards establishing a far better evidence base to use to make these decisions.

References

Arvelius, P., Strandberg, E., Fiske, W.F. (2014) The Swedish Armed Forces temperament test gives information on genetic differences among dogs. *Journal of Veterinary Behavior: Clinical Applications and Research.* 9, (6), 281–289.

Bell, A.M., Hankison, S.J., Laskowski, K.L. (2009) The repeatability of behavior: A meta-analysis. *Animal Behavior.* 77, (4), 771–784.

Butt, T. (2004) Personality theories: Trait, biological and cognitive social approaches, *Understanding People*, Basingstoke and New York, Palgrave MacMillan.

Diederich, C., Giffroy, J.M. (2006) Behavioural testing in dogs: A review of methodology in search for standardisation. *Applied Animal Behaviour Science.* 97, (1), 51–73.

Dowling-Guyer, S., Marder, A., D'Arpino, S. (2011) Behavioral traits detected in shelter dogs by a behavior evaluation. *Applied Animal Behaviour Science.* 130, (3-4), 107–114.

Duffy, D.L., Serpell, J.A. (2012) Predictive validity of a method for evaluating temperament in young guide and service dogs. *Applied Animal Behaviour Science.* 138, (1-2), 99–109.

Fratkin, J.L., Sinn, D.L., Patall, E.A., Gosling, S.D. (2013) Personality consistency in dogs: A meta-analysis. *PLoS ONE*. 8, (1), e54907.

Harvey, N.D., Craigon, P.J., Sommerville, R., McMillan, C., Green, M., England, G.C.W., Asher, L. (2016) Test-retest reliability and predictive validity of a juvenile guide dog behavior test. *Journal of Veterinary Behavior: Clinical Applications and Research. 11,* (2), 65-77.

Jones, A.C., Gosling, S.D. (2005) Temperament and personality in dogs (*Canis familiaris*): A review and evaluation of past research. *Applied Animal Behaviour Science.* 95, (1-2), 1–53.

Mornement, K.M., Coleman, G.J., Toukhsati, S., Bennett, P.C. (2014) Development of the behavioural assessment for re-homing K9's (B.A.R.K.) protocol. *Applied Animal Behaviour Science.* 151, 75–83.

Rayment, D.J., De Groef, B., Peters, R.A., Marston, L.C. (2015) Applied personality assessment in domestic dogs: Limitations and caveats. *Applied Animal Behaviour Science.* 163, 1–19.

Riemer, S., Muller, C., Viranyi, Z., Huber, L., Range, F. (2014) The predictive value of early behavioral assessments in pet dogs—A longitudinal study from neonates to adults. *PLoS ONE.* 9, e101237.

Sinn, D.L., Gosling, S.D., Hilliard, S. (2010) Personality and performance in military working dogs: Reliability and predictive validity of behavioral tests. *Applied Animal Behaviour Science.* 127, (1-2), 51–66.

Taylor, K.D., Mills, D.S. (2006). The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, 1, (3), 94-108.

Valsecchi, P., Barnard, S., Stefanni, C., Normando, S. (2011) Temperament test for re-homed dogs validated through direct behavioral observation in shelter and home environment. *Journal of Veterinary Behavior: Clinical Applications and Research*. 6, (3), 161–177.

Wiener, P., Haskell, M.J.. (2016) Use of questionnaire-based data to assess dog personality. *Journal of Veterinary Behavior: Clinical Applications and Research*. 16, 81-85.

Wilsson, E., Sinn, D.L. (2012) Are there differences between behavioral measurement methods? A comparison of the predictive validity of two ratings methods in a working dog program. *Applied Animal Behaviour Science.* (141), 158–172.