Behavioral testing in puppies as working dogs: a mini-review

Jara Gutiérrez Rufo*

Dipartimento Scienze Veterinarie, Università di Pisa

Abstract: This review will focus on the behavioral study of puppies with the aim of determining their characteristics as suitable working dogs. In search of a standardisation, some examples of material and methodological aspects of behavioral testing in puppies have been also collected.

In puppy testing, some authors keen to follow their behavioral development by retesting the same animals at various periods of their lives or by attempting to predict what characteristics the dog will have in adulthood and what task it will be most suited to. Not surprisingly, given the variety of fields doing research on dog puppies, the studies varied widely in their aims.

These principal goals include selecting the right puppy for the right family, selecting dogs suitable as guide or therapy dogs, social skills, exploratory behavior, genetic influence on dog behavior, assessing the temperament of dogs in rescue shelters, measuring adaptive behavior.

Behavior of puppies changes rapidly, so when it is tested, it is not very likely to find a correspondence to adult behavior. It is important to test dogs later, when their behavior has matured and when changes per time are less, in order for the results to predict the adult phenotype. Individual variation in behavior observed among puppies can be explained both by hereditary factors and by effects of common litter environment. From this point of view, breeding programs aimed to improve behavior in dogs may not be based on information collected on tests performed as early as at eight weeks of age.

Key Words: puppy, behavior, test, working dog.

* Correspondence Author: jaragutierrez@hotmail.es

Introduction

In puppy testing, some authors keen to follow their behavioral development by retesting the same animals at various periods of their lives (Fox & Spencer, 1969; Wright, 1980) or by attempting to predict what characteristics the dog will have in adulthood, and what task it will be most suited to (Diederich & Giffroy, 2006). Not surprisingly, given the variety of fields doing research on dog puppies, the studies varied widely in their aims.

These principal goals include selecting the right puppy for the right family (Campbell, 1975; Fogle, 1990; Van Den Borg et al., 1991; Bondarenko, 1995; Coren, 1995), selecting dogs suitable as guide or therapy dogs (Pfaffenberger, 1963; Pfaffenberger and Scott, 1976; Goddard & Beilharz, 1984a,b; Schaffer & Phillips, 1993; Murphy, 1995), social skills (Hare et al., 2002; Topal et al., 2005; Howell & Bennet, 2011), exploratory behavior (Thompson & Heron, 1954; Plutchik, 1971; Fox and Spencer, 1969; Fox, 1972; Hughes, 1972; Fox & Andrews, 1973; Walsh & Cummins, 1976; Henderson, 1980; Diederich & Giffroy, 2006), genetic influence on dog behavior (Reuterwall & Ryman, 1973; Greyvenstein, 1982), assessing the temperament of dogs in rescue shelters (Ledger & Baxter, 1997), measuring adaptive behavior (Coren, 1995). Other studies on behavioral tests
have had the purpose of investigating if behavior tests could be used to select service dogs for various kinds of work and breeding (Wilsson & Sundgren, 1997a), predicting the success of police (Reuterwall & Ryman, 1973; Fält et al., 1982; Verga, 1983; Fält, 1984; Mackenzie et al., 1985) and military dogs (Rooney et al., 2003).

This review will focus on the behavioral study of puppies with the aim of determining their characteristics as suitable working dogs. In search of a standardisation, some examples of material and methodological aspects of behavioral testing in puppies have been also collected.

The goal of Puppy Aptitude Tests (PAT) is to help people to select a suitable breed and dog. Two types of assessment were reported here: breed profile created from Expert Ratings of Breed (e.g., Coren, 1995; Hart & Miller, 1985; Hart & Hart, 1985), and puppy-behavior test, a type of Test Battery, to be performed by the puppy buyer (e.g., Campbell, 1972; examined in Beaudet et al., 1994; Campbell 1975). Beaudet et al. (1994) demonstrated a predictive value of the ‘Tests of Campbell’ between 7 and 16 weeks of age, by introducing into the analysis an independent supplementary measure (total number of squares crossed in the test arena). Even though these tests are well-known, their predictive value has been demonstrated neither by their author nor by Young (1988). Additionally, Lindsay (2000) suggested that puppy temperament tests should not be used to predict adult efficiency or the potential exhibition of adult behavior patterns but should be used to identify puppy’s strengths and weaknesses at the time of testing.

Despite also Jones & Gosling (2005) stated, in one of the most recent review on temperament and personality in dogs, that “tests of young puppies are not valid predictors of their future behavior”, today the opinions on predictive value of PAT are still controversial (Table 1).

Table 1. Examples of controversy about the efficacy in the predictability of behavior in dogs using PAT.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Specific topic of the research</th>
<th>Does puppy testing show a correlation with adult behavior?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martínek et al. (1970, 1975)</td>
<td>Habituation and learning in puppies - Habituation and learning in adult dogs</td>
<td>Yes</td>
</tr>
<tr>
<td>Scott &amp; Bielfelt (1976)</td>
<td>Temperament-based selective breeding - Selection of guide adult dogs</td>
<td>No</td>
</tr>
<tr>
<td>Goddard &amp; Beilharz (1986)</td>
<td>Prediction of fearfulness at 6 months of age - Disqualifying guide dogs in adulthood</td>
<td>Yes</td>
</tr>
<tr>
<td>Schaffer &amp; Phillips (1993) and Bondarenko (1995)</td>
<td>Predictive success of the puppy to become a successful police dog</td>
<td>No</td>
</tr>
<tr>
<td>Wilsson &amp; Sundgren (1998)</td>
<td>Puppies exploration - Adult dog’s ability to cooperate</td>
<td>Yes</td>
</tr>
<tr>
<td>Slabbert &amp; Odendaal (1999)</td>
<td>Retrieval test at 2 months and aggression test at 9 months - Success as an adult police dog</td>
<td>Yes</td>
</tr>
<tr>
<td>Rooney et al. (2003)</td>
<td>Behavioral activity of 7-week-old German shepherd puppies - Police efficiency during adulthood</td>
<td>Yes (German shepherd puppies)</td>
</tr>
<tr>
<td></td>
<td>Validity for specialist search work of 8-week-old Labrador retriever puppies - Retesting validity for specialist search work at 11 months of age</td>
<td>No (Labrador retriever puppies)</td>
</tr>
<tr>
<td>Scott &amp; Bielfelt (1976)</td>
<td>Repetitive tests performed on puppies 8-12 weeks of age - Prediction of their future usefulness as guide dogs</td>
<td>Yes</td>
</tr>
<tr>
<td>Svobodová et al. (2007)</td>
<td>Puppy test as a predictor of later behavior and trainability as police and military dogs</td>
<td>No</td>
</tr>
</tbody>
</table>
Since the time when Clarence J. Pfaffenberger with his co-workers started testing puppies selected to become guide dogs (Pfaffenberger, 1963; Lambin, 1994), PAT have become a frequent tool for choosing individuals suitable for specific tasks (Slabbert & Odendaal, 1999) or just as household pets (Campbell, 1972). Pfaffenberger (1963) tested puppies to reveal their inherited character traits which would prevent them from becoming guide dogs. Tests were done at an early age, during a sensitive period of learning, between 6 and 16 weeks and they consisted of reactions to various startling and tactile stimuli, novel situations and willingness to fetch. Results showed a high correlation between dogs who performed well the tests at an early age and those who became guide dogs.

Later, Michael Fox’s work, on canine behavior, popularised the notion of predictability in puppy behavior and by the 1970s PAT became familiar and a frequently recommended technique for selecting suitable dogs (Fox, 1970).

Noteworthy results were achieved by Martínek et al. (1970, 1975) too, who tested German shepherd pups’ rate of habituation and learning as early as at the age of 2 and 4 months and compared the results with the performance of the same individuals at the age of 1 year. The results achieved from PAT were in good correlation with the tests performed on 1-year-old dogs. This suggests that at least some behavioral activity of a puppy may be a good predictor of the dog’s behavior in adulthood.

In other study, the number of objects that a puppy investigated, in a given period of time, was a predictor of the adult dog’s ability to cooperate (Wilsson & Sundgren, 1998).

Rooney et al. (2003) presented a study examining the validity of PAT for dogs selected for specialist search work. The test was performed on Labrador retrievers, 8 weeks old, and the results compared with the retest at 11 months of age. The authors concluded PAT, at least in Labrador retrievers, are unlikely to be useful predictors of adult behavior. But when behavior of 7-week-old German shepherd puppies was tested for predicting police efficiency of the individual, test seemed to be an appropriate tool for predicting suitable dogs as early as 7 weeks of age.

Several other studies suggested PAT could predict adult efficiency and behavioral trait reliably. Fox (1972) and Campbell (1975) have proposed behavioral tests used to select puppies from a litter. Nonetheless, neither of these tests were evaluated or proven to have any predictive value (Wilsson et al., 1998).

Slabbert & Odendaal (1999) proposed a set of tests for the South African Police Service (SAPS). This test battery included crossing obstacle courses, retrieving objects, novel and startling visual and auditory stimuli, and situations attempting to induce aggressive behavior. High scores on the retrieval test at 2 months and the aggression test at 9 months significantly predicted success as an adult police dog. A high percentage of puppies which had low scores for the obstacle test, did not become police dogs (n=96; p<0.005). A moderate percentage of puppies which achieved high scores for this test became successful police dogs (n=71; p<0.005). A comparison was made between the percentage of puppies which achieved high scores and became police dogs in adulthood, and those which achieved low scores and failed to become police dogs (Fig. 1).

A high percentage (78.1%) of puppies which achieved low scores at the aggression test did not become police dogs (p<0.001), while a fairly high percentage (69%) of those which achieved high scores went on to become police dogs.
during adulthood. Startle test scores of individual 16-week old puppies \((n=167)\). Comparison was made between the percentage of juveniles which achieved high scores and went on to become police dogs in adulthood, and those which achieved low scores and failed to become police dogs (Fig. 2).

Fearfulness, characteristic often disqualifying guide dogs, may be predicted, according Goddard & Beilharz (1984a, 1986), as early as at 3 months of age, but the results are more accurate if dogs are tested at the age of 6 months.

After, they modified somewhat this statement and it was suggested that the predictive value increased if dogs were tested later in life (Goddard & Beilharz, 1986). What is more, they reported some correlations between scores at the various ages at which they assessed guide-dog puppies, making the point that the correlations increased as the dogs aged. Their purpose was to evaluate this change and to determine how old a puppy must be for the puppy’s fearfulness level to predict its adult fearfulness (Goddard & Beilharz, 1984a, b). They studied correlations between measurements of responses, such as approach avoidance, nature of contact with stimulus, tail position and posture, to many stimuli. They finally reported that adult performance could be predicted to some degree from fearfulness at 3 months of age, but that the accuracy of the prediction improved with age.

Traits related to fearfulness and confidence have been also found to be stable from early age in dogs (Royce, 1955; Pfaffenberger et al., 1976; Goddard & Beilharz, 1986; Slabbert & Odendaal, 1999).

The effects of this temperament-based selective breeding can be seen in various programs. For example, selective breeding based on puppy test performance scores at the Guide Dogs for the Blind training center in San Rafael, California (Scott & Bielfelt, 1976), lead to an improvement in puppy test scores over successive generations; which scores was not matched in the rates at which adult dogs became successful guide dogs, suggesting the puppy tests may not be an ideal criterion for selecting guide dogs, at least in this high-functioning group of subjects.

On the other hand, PAT were not considered reliable, by Schaffer & Phillips (1993) and Bondarenko (1995), in predicting whether a puppy would become a successful police dog.

A disproportionately considerable number of the dogs examined were animals specially bred and trained for definite working programs. This is a key point to take into account by people seeking to use the research on temperament to understand and predict the behavior of pet or shelter dogs (Jones & Gosling, 2005).

However, Jones & Gosling (2005) found that the scores in these tests were not appropriate at all for their analyses because they sought to index change in the subjects, not repeatability of the test.

Scott & Bielfelt (1976) described several repetitive tests performed on puppies 8-12 weeks of age with the purpose to predict their future usefulness as guide dogs. Tests were acceptably predictive, although heritability of the tested characteristics was found to be low and effects of maternal environment to be high.

To investigate if a puppy test could be used to predict adult usefulness, the hereditary effects on the behavior of puppies can be estimated. Willson & Sundgren (1997a) compared puppy tests to the results at adult age to select dogs for work and breeding. 1235 eight-week old German shepherd puppies were tested during the years 1978-1983. Most of these dogs were also tested as adults,
450-600 days of age. The test performed on adult dogs also included situations where the dog was threatened or exposed to fear eliciting stimuli, situations not appropriate for puppies. Therefore, the intention was not to adapt the puppy test to the adult test. The puppy test focused instead on characteristics with considerably variability among eight-week old puppies, i.e. sociability, independence, fearfulness, competitiveness, general activity and explorative behavior. Some of the test situations were adapted from earlier tests (Scott & Fuller, 1965; Fox & Stelzner, 1966; Fox, 1972; Scott & Bielfelt, 1976; Fuller & Clark, 1968). The test is divided into parts (1) Vocalisation when isolated and (2) Arena test.

Part (1) included yelp, shriek, contact to the person, willingness to retrieve objects – valid to measure cooperativeness; it has previously been shown to have a high predictive value (Scott & Bielfelt, 1976) –, fetch, retrieve the ball, large ball - to elicit a fear response – and tug of war – to evaluate the puppy’s future interest in objects –. In Part (2) all puppies were moved to an unfamiliar building. The floor was marked in squares, 0.6 x 0.6 m. In the middle of the arena a circle 2 m in diameter was marked. Out on the floor of the arena are four objects; a piece of wood and three objects used in part (1) – the tennis ball, the large ball and the cotton rag –. The purpose of this part is to test the dog’s explorative behavior with a passive person present while measuring time spending inside the circle, activity – number of squares the puppy moves over – and number of objects visited during the test (Wilsson & Sundgren, 1998).

They concluded that correspondence of 8-week old puppy test results to performance at adult age was negligible and the puppy test therefore was found useless as a predictor of later behavior and trainability as police and military dogs (Svobodová et al., 2007). Moreover, results showed scarcely behavioral sex differences in dogs (regression coefficient (r) = -5.66; p<0.001, Table 2), whereas Scott and Fuller (1965) had found males to be dominant over females in competitive situations at the age of eight weeks.

It is important to consider that puppies are maturing rapidly at the age of eight weeks so, level of maturation can cause a significant impact on the results (Wilsson & Sundgren, 1998). If the behavior of puppies is affected by degree of maturation, this would certainly diminish the predictive value of the test. Considering this, one would expect the predictive value of a test on young dogs to increase if dogs are tested later when the differences in behavior is less affected by degree of maturation. Medium high to high estimates of heritabilities found in puppies (Table 3; Wilsson and Sundgren, 1998) and on adult dogs (Wilsson & Sundgren, 1997b) show that behavior is strongly affected by hereditary factors.

Table 2. LS-means in the puppy test results for each sex respectively. Differences between sexes and level of significance (Wilsson & Sundgren, 1998).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean (Males n = 320)</th>
<th>Mean (Females n = 310)</th>
<th>Difference (m – f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yelp</td>
<td>44.3</td>
<td>50.8</td>
<td>-6.5</td>
</tr>
<tr>
<td>Shriek</td>
<td>222.4</td>
<td>226.6</td>
<td>-4.2</td>
</tr>
<tr>
<td>Contact I</td>
<td>4.8</td>
<td>4.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Fetch</td>
<td>61.6</td>
<td>49.4</td>
<td>12.2*</td>
</tr>
<tr>
<td>Retrieve</td>
<td>4.1</td>
<td>4.2</td>
<td>-0.1</td>
</tr>
<tr>
<td>Large ball</td>
<td>4.4</td>
<td>4.6</td>
<td>-0.2</td>
</tr>
<tr>
<td>Tug of war</td>
<td>3.3</td>
<td>3.3</td>
<td>0.0</td>
</tr>
<tr>
<td>Activity</td>
<td>84.8</td>
<td>92.3</td>
<td>-7.5*</td>
</tr>
<tr>
<td>Contact II</td>
<td>113.8</td>
<td>94.0</td>
<td>19.8**</td>
</tr>
<tr>
<td>Objects visited</td>
<td>6.9</td>
<td>8.1</td>
<td>-1.1***</td>
</tr>
</tbody>
</table>

Level of significance: *p < 0.05; **p < 0.01; ***p < 0.001.
The lack of correspondence between puppy and adult test results could potentially be explained if juvenile behavior is governed by different genes than behavior at adult age (Wilsson & Sundgren, 1998). Using a combination of temperament assessments, Burghardt (2003) concluded that the predictive value of tests rapidly increases until at least 6 months of age in Belgian Malinois Dogs.

To sum up, behavior of puppies changes rapidly, so when their behavior is tested, it is not very likely to find a correspondence to adult behavior. It is important to test dogs later, when their behavior has matured and when changes per time are less, for results able to predict the adult phenotype (Goddard & Beilharz, 1986; Wilsson & Sundgren, 1998). Individual variation in behavior observed among puppies can be explained both by hereditary factors and by effects of common litter environment. From this point of view, breeding programs aimed to improve behavior in dogs may not be based on information collected on tests performed as early as at eight weeks of age (Wilsson & Sundgren, 1998).

Socialisation and training programs may be useful as a starting point for assessing possible problematic behavior in puppies and are effective in producing well trained working dogs (Seksel et al., 1999). Previous studies such as those by Clarke et al. (1951), Thomson and Heron (1954), Melzack, (1968), Fuller (1964), compared dogs that had virtually no socialisation at all with those that had normal levels of socialisation. The experiments were carried out under laboratory conditions, not the home environment. The data appear to suggest that less, but not more than normal levels, will affect later behavior. It may be possible that a minimum level of social interaction is needed at an early age, but any above that level (extra socialisation) does not necessarily increase social behavior.

This suggestion is supported by the findings of Pfaffengerber et al. (1976) who found that puppies taken out of their kennel environment prior to 12 weeks of age were more likely to succeed as guide dogs. Wolle (1990) cited in Serpell (1995) found that laboratory beagles needed less than 5 min a week of social contact with humans to be 'adequately socialised'. Other species can be used as interspecific social stimuli. Puppies may be reared with cats and then presented to dog littermates to study their socialisation to their own species (Fox, 1971) and their social behavior to a cat in a wire cage have been also rated in Seksel et al. (1999).

Table 3. Heritabilities ± S.E. estimated on the puppy test results for 277 male and 277 female puppies tested year 3-6. The highest estimates (combined values) were found on the score groups tug of war, activity, and contact II. Other estimates of heritability ranges from 0.20-0.27. Higher estimates on dam than on sire were found in the score groups shriek, contact I, tug of war, activity, contact II and objects visited (Wilsson & Sundgren, 1998).

<table>
<thead>
<tr>
<th>Score Group</th>
<th>Sire</th>
<th>Dam</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yelp</td>
<td>0.20 ± 0.24</td>
<td>0.25 ± 0.25</td>
<td>0.22 ± 0.08</td>
</tr>
<tr>
<td>Shriek</td>
<td>0.24 ± 0.33</td>
<td>0.33 ± 0.73</td>
<td>0.24 ± 0.09</td>
</tr>
<tr>
<td>Contact I</td>
<td>0.05 ± 0.26</td>
<td>0.36 ± 0.28</td>
<td>0.21 ± 0.08</td>
</tr>
<tr>
<td>Fetch</td>
<td>0.22 ± 0.23</td>
<td>0.19 ± 0.24</td>
<td>0.21 ± 0.08</td>
</tr>
<tr>
<td>Retrieve</td>
<td>0.23 ± 0.23</td>
<td>0.17 ± 0.23</td>
<td>0.20 ± 0.08</td>
</tr>
<tr>
<td>Large ball</td>
<td>0.27 ± 0.25</td>
<td>0.27 ± 0.25</td>
<td>0.27 ± 0.09</td>
</tr>
<tr>
<td>Tug of war</td>
<td>0.21 ± 0.34</td>
<td>0.76 ± 0.36</td>
<td>0.48 ± 0.11</td>
</tr>
<tr>
<td>Activity</td>
<td>−0.60 ± 0.51</td>
<td>1.67 ± 0.60</td>
<td>0.53 ± 0.13</td>
</tr>
<tr>
<td>Contact II</td>
<td>0.04 ± 0.35</td>
<td>0.80 ± 0.38</td>
<td>0.42 ± 0.10</td>
</tr>
<tr>
<td>Objects visited</td>
<td>−0.28 ± 0.34</td>
<td>0.82 ± 0.40</td>
<td>0.27 ± 0.09</td>
</tr>
</tbody>
</table>
References


Campbell W.E. Behavior problems in dogs. Behavior problems in dogs. 1975


Martinek Z., Lat J., Hartl K. About the possibility of predicting the performance of adult guard dogs from early behavior. In: Symposium on Basic Mechanisms of Learning and Memory, Marianske Lazne, 1970


Schaffer C.B., Phillips J. The Tuskagee Behavior Test for selecting Therapy Dogs. Tuskagee University School of Veterinary Medicine, Tuskagee. 1993


I test comportamentali nei cuccioli per la scelta di cani da lavoro: una mini-review

Jara Gutiérrez Rufo

Dipartimento Scienze Veterinarie, Università di Pisa

Sintesi

Questa review ha come argomento i test comportamentali effettuati sui cuccioli con lo scopo di determinare le loro caratteristiche comportamentali necessarie per far di loro dei cani da lavoro.

Nei test per i cuccioli alcuni autori hanno seguito i soggetti, testandoli a diverse età, cercando di predire le caratteristiche comportamentali del cane adulto e quale compito lavorativo esso potrà svolgere.

Le principali finalità di queste ricerche sono selezionare i cuccioli adatti come animali da famiglia o come cani guida per non vedenti o da assistenza, dotati di particolari abilità sociali e comunicative, oppure valutare il comportamento di cani ospitati in canili.

Il comportamento dei cuccioli cambia rapidamente, così quando esso è valutato, non è facile trovare una corrispondenza col comportamento dell’adulto. È quindi importante testare il cane in età più avanzata, quando il suo comportamento è maturato e quando i cambiamenti sono minori nell’arco di tempo considerato, in modo che i risultati possano predire più fedelmente il fenotipo da adulto.

Le differenze individuali nel comportamento, osservate tra i cuccioli, possono essere spiegate sia attribuendole a fattori ereditari sia agli effetti dell’ambiente sulla cucciola.

Da questo punto di vista, programmi di allevamento finalizzati a migliorare il comportamento nei cani non possono essere basate su informazioni raccolte sui test eseguiti a otto settimane di età.