# Genetic Evaluation of Behaviour in Dogs

## Per Arvelius

*Faculty of Veterinary Medicine and Animal Science*
*Department of Animal Breeding and Genetics*
*Uppsala*

Doctoral Thesis
Swedish University of Agricultural Sciences
Uppsala 2014

Cover: Ebba
    (photo: Malin Wijk)

# Genetic Evaluation of Behaviour in Dogs

## Abstract

A dog's behavioural characteristics are important for the dog, for the dog owner and for society as a whole. Behavioural traits can be changed by breeding, but to be effective when selecting breeding animals, good methods for measuring behaviour are essential. The aim of this thesis was to provide information on a number of dog behavioural measurement methods regarding their potential to be used for genetic evaluation: the Herding Trait Characterisation, the Swedish and Norwegian English Setter field trials, the Swedish Armed Forces temperament test, the Dog Mentality Assessment (DMA), and an extended version of the Canine Behavioral Assessment and Research Questionnaire. The aim was also to advance our understanding of factors affecting the usefulness of behavioural measurements for breeding purposes.

Average heritabilities for behavioural variables (items) within measurement method ranged from 0.1 to 0.3, and the items were markedly influenced by systematic environmental effects. All studied measurement methods can be used for selection of breeding animals, but selection based on individual performance is suboptimal. Using BLUP breeding values would substantially increase accuracy of selection and the potential genetic progress and is therefore recommended. Rough Collie results from DMA showed strong genetic correlations with important everyday life traits as described by dog owners in the questionnaire. Therefore, in order to improve everyday life behaviour in Rough Collie, DMA breeding values for relevant traits should be used for selection.

The results indicated that from a heritability perspective, behavioural measurements should be objective rather than subjective, and neutral rather than passing value judgments. Collaboration between countries within breed is also advised because a joint genetic evaluation increases the number of selection candidates, and may also increase breeding value accuracies rather dramatically, as was shown for the English Setter field trial results from Sweden and Norway.

For half of the studied methods, the measured items were summarized into composite traits. Heritability estimates for composite traits were higher than the average of the items used for creating these traits. Because the composite traits also can be expected to be more stable over time and between situations, it would be advisable to use them as selection traits.

*Keywords:* behaviour, breeding, dog, genetic evaluation, genetic progress, herding trait, heritability, hunting trait, selection, temperament

*Author's address:* Per Arvelius, SLU, Department of Animal Breeding and Genetics, P.O. Box 7023, 750 07 Uppsala, Sweden
*E-mail:* Per.Arvelius@ slu.se

*Ibland oförnuftige kreatur äger hunden onekligen de snilles gåfwor, som närmast likna förnuft. Han är mera läracktig, än något annat djur, han kan läras gå ut då naturen kräfwer, genom sin waksamhet bewara hus och gård från främmande kreatur, förstår til en del hwad man teknar och talar med honom, gifwer tillkjänna när något owanligit eller fara är å färde, om tjufwar eller andra obekanta människjor annalkas: Han kan ock skilja främmande kreatur från sin Husbondas, m. m.*

Carl von Linné

# Contents

# List of Publications

This thesis is based on the work contained in the following papers, referred to by Roman numerals in the text:

I   Arvelius, P., Malm, S., Svartberg, K. and Strandberg, E. (2013). Measuring herding behavior in Border collie – effect of protocol structure on usefulness for selection. *Journal of Veterinary Behavior: Clinical Applications and Research* 8, 9-18.

II  Arvelius, P. and Klemetsdal, G. (2013). How Swedish breeders can substantially increase the genetic gain for the English Setter's hunting traits. *Journal of Animal Breeding and Genetics* 130, 142-153.

III Arvelius, P., Strandberg, E. and Fikse, W.F. (2014). The Swedish Armed Forces temperament test gives information on genetic differences among dogs. *Journal of Veterinary Behavior: Clinical Applications and Research* DOI: http://dx.doi.org/10.1016/j.jveb.2014.06.008.

IV  Arvelius, P., Eken Asp, H., Fikse, W.F., Strandberg, E. and Nilsson, K. (2014). Genetic analysis of a temperament test as a tool to select against everyday life fearfulness in Rough Collie. *Journal of Animal Science* 92, 4843-4855.

The papers are reproduced with the permission of the publishers.

# Abbreviations

BLUP      best linear unbiased prediction
BR      behavioural rating
C-BARQ      Canine Behavioral Assessment and Research Questionnaire
DMA      Dog Mentality Assessment
EBV      estimated breeding value
ES FT      English Setter field trial
FS      factor score
HTC      Herding Trait Characterisation
SAF TT      Swedish Armed Forces temperament test
SR      subjective rating
SS      summated scale

# 1  Introduction

A dog's behavioural characteristics are important for the dog, for the dog owner and for society as a whole. Because behavioural traits are heritable and can be changed by breeding, they should be included as an important part of the breeding goal. Within a dog breed, the breeders – each of them usually with only a small production of puppies – relatively independently select which dogs to use for breeding. The breeding goal may differ between breeders and thus genetic progress is often slow for many traits.

In livestock breeding, effective methods for evaluating animals genetically, and to select which individuals to use for breeding, have been developed and extensively used for several decades with great success. Instead of selecting breeding animals based on their phenotypic performance, these methods allow for estimating an animal's breeding value by adjusting the phenotype for environmental factors and by taking information on relatives into account. This makes it possible to be more accurate in selecting breeding animals for the genetic qualities they will contribute to the offspring. In dog breeding, however, people still today mainly practice phenotypic selection.

The overall aim of this thesis was to investigate the prospects for improving dog behaviour by breeding. Dog breeders would potentially benefit substantially in terms of faster genetic progress for important behavioural traits, if modern methods for genetic evaluation were applied. For this to function well, it is essential to have good methods for measuring the traits of interest. In this thesis, a number of dog behavioural measurement methods were evaluated for their potential to be used for genetic evaluation, and with the purpose of advancing our understanding of factors affecting the usefulness of behavioural measurements for breeding purposes.

# 2   Background

## 2.1   Breeding

Animal breeding is about choosing the genetically best individuals as parents, and thereby bring about genetically improved offspring generations. The structure of a breeding program can differ between species and populations. But even if there are differences in, for example, how systematic or advanced programs are, or how much emphasis is put on the various parts, the basic components can be expected to be more or less the same (Figure 1). There is a breeding goal, towards which the breeders strive to change the population of animals. To do this, animal phenotypes of traits of interest for the breeding goal are recorded. The animals available for breeding are ranked with the purpose to identify the genetically best individuals and the animals to be used as parents are selected and mated to produce the next generation. Then the process starts again with recording the phenotypes of the offspring and so on. If successful, this will lead to a genetic progress, meaning that each offspring generation becomes genetically better than its parent generation. Usually, the intention is to achieve the genetic progress with a limited increase in inbreeding level.
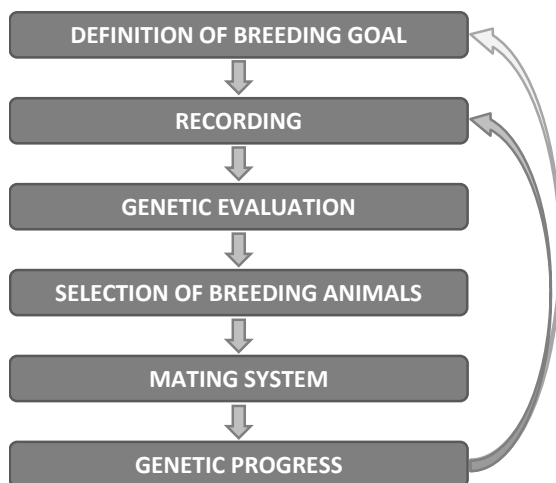


DEFINITION OF BREEDING GOAL

RECORDING

GENETIC EVALUATION

SELECTION OF BREEDING ANIMALS

MATING SYSTEM

GENETIC PROGRESS

*Figure 1*. Principles of a breeding program.

## 2.1.1 Breeding goal

To define the breeding goal, it must be decided what traits are important, and also which are more important than others in order to know how much emphasis to put on each trait. One pitfall is to put too much emphasis on traits only because they are easy to measure, and also the opposite, to underestimate the value of important traits just because they are difficult to measure. If, for example, striving to breed for better working dogs, it might not help much to include size of the ears among the selection traits. If this trait were included, the genetic progress for the traits that really are important for working ability would most likely become slower. In a worst case scenario, ear size will turn out to be unfavourably correlated to, for example, important health traits. One illustrative example comes from the breeding for milk production in cows, where selection in the U.S. used to be partly based on the morphological trait Dairy form. The problem was that this lead to a slower genetic progress for the breeding goal traits – milk yield and disease resistance – compared to if selection had been based on milk yield alone (Rogers *et al.*, 1999).

## 2.1.2 Recording

To be able to breed systematically, traits in the breeding goal should be measured and recorded for as many animals as possible. If, for example, 10 new breeding animals are to be selected, and selection is on phenotype, selection can be much more intensive – potentially generating a faster genetic progress – if records exist for 1000 animals, compared to if only 100 animals have been recorded. It also becomes easier to find more unrelated animals, and thus to keep a lower inbreeding rate.

In a modern breeding program, ranking of a potential breeding animal is not based only on the animal's own result. To more accurately identify the genetically best individuals for a trait, results from relatives are also taken into account when estimating an animal's so-called breeding value (EBV) for the trait. This emphasizes even more the need for extensive recording of phenotypes, but also highlights the need for correct pedigree data. To improve the accuracy of the EBVs further, various environmental factors can be adjusted for. Examples of such factors, which might affect a dog's behaviour, are the time of the year or age of the animal when the registration is done. To be able to adjust for environmental factors, they have to be recorded.

In summary, the more animals for which records exist, containing phenotypes as well as information on environmental factors, and the more complete the pedigree data, the greater possibilities to breed successfully.

### 2.1.3 Genetic evaluation, selection and mating

Based on the breeding goal and the measurements for relevant traits, the animals are ranked with the purpose to identify the genetically best individuals, and the animals to be used as parents are selected. To do this, various methods can be applied. Within dog breeding, animals are still today usually ranked entirely based on their phenotypic performance. In livestock breeding more advanced methods have been used for a long time, taking information on relatives' performance and environmental factors into account when evaluating animals for quantitative traits. Quantitative traits are under the influence of many genes together with non-genetic causes (environmental factors). Most traits of economic importance to livestock breeders are quantitative (Falconer and Mackay, 1996), and so are most dog behavioural traits (van Rooy *et al.*, 2014). Quantitative traits typically are not either-or, instead differences are gradual. For example, a dog is not aggressive or non-aggressive, it is more aggressive or less aggressive. Information on an individual's phenotype for a trait with quantitative background does not automatically mean that the genotype can be easily described. The reason is – as already mentioned – that the phenotype is not only influenced by genetic factors. The environment too, in which also any measurement error is included, affects the phenotype. The statistical measure heritability describes how much of the measured variance in a trait has additive genetic background. Expressed in another way, the heritability describes how similar relatives are, and the higher the heritability, the faster a trait can be changed by breeding.

The BLUP method (Best Linear Unbiased Prediction) makes it possible to systematically evaluate animals from a breeding perspective for traits with quantitative background. By including phenotypic information also on relatives and adjusting for systematic environmental effects to estimate animals' breeding values, accuracy of selection is improved, generating increased genetic progress. Henderson (1976) showed how to handle pedigree information in a way that made it possible to include all pedigree information also in very large populations when estimating breeding values. Previously, this had been a limitation. By taking all relatives with phenotypic information into account when estimating an individual's breeding value, accuracy increased even more.

The fact that information from relatives is used when estimating an individual's BLUP breeding value does not only have positive consequences in terms of increased accuracy. It also has the effect that close relatives tend to get similar breeding values. This increases the probability of selecting close relatives to become parents of the next generation which, in turn, would increase the inbreeding rate. Optimum contribution selection (Meuwissen,

1997) addressed this dilemma by selecting which animals that should be used for breeding to generate the fastest genetic progress (based on, for example, BLUP breeding values) given a predefined constraint on what inbreeding rate can be tolerated.

Currently, animal breeding is going through a revolution due to the possibility to select animals based on genomic information (Calus, 2010). The basic principle for genomic selection is that the effect on a trait of a large number of DNA markers is estimated by combining genomic information for the animals in a reference population with phenotypic information from the same animals. Information on the marker effects can then be used to estimate genomic breeding values for genotyped selection candidates. A main advantage with genomic selection is that animals can be accurately selected without having their phenotypes recorded (Hayes *et al.*, 2013). This may be beneficial when it is difficult or expensive to make phenotypic measurements, or if the trait is expressed late in life.

After having selected the animals to be used for breeding, there might be reason to have some sort of system for how to mate them. For example, a strategy might be chosen where weaknesses in one parent are compensated by choosing a mate that is good for corresponding traits. Other examples could be making sure not to mate animals carrying the same allele for a serious recessive disease, or close relatives, with each other.

### 2.1.4  Genetic progress

The ultimate goal in a breeding program is genetic progress – that the animals become genetically better with each generation. Because running a breeding program is a long-term operation, it is also essential to keep the inbreeding rate as low as possible to decrease the risk for recessive diseases to be expressed, and to avoid inbreeding depression. With time, inbreeding also leads to lower genetic variation in general, which in turn leads to slower genetic progress.

## 2.2   Dog behaviour and breeding

### 2.2.1  Dog behaviour is important

The dog was the first domesticated animal. Dogs were present in Europe and the Far East about 15,000 years ago, but when and where domestication took place – and if it happened at several separate occasions or only once – is unclear (Larson *et al.*, 2012; Larson and Bradley, 2014). The dog's closest relative among wild species living today, is the grey wolf (Clutton-Brock, 1995; Lindblad-Toh, 2005). Domestication is considered to have generated behavioural changes, such as reduced fearfulness and aggressiveness

towards humans (Serpell and Duffy, 2014), and regarding sociability and cognition (Marshall-Pescini and Kaminski, 2014). Axelsson *et al.* (2013) identified 19 regions in the genome that were believed to have been under selection as an effect of domestication, and that contained genes of relevance to brain function.

Serpell and Duffy (2014) noted that what makes the dog unique compared to other domesticated species is that dog selection has primarily been on behaviour. Dogs have been appreciated for different working and social abilities, for example guarding, hunting, fighting or as pets, while livestock selection has focused on production traits, such as egg production and growth rate. In the mid-19th century, dog breeding started to become formalized in the way that dog breeds were created; breed clubs were formed and so-called breed standards written, seemingly emphasizing appearance over functionality (McGreevy and Nicholas, 1999; Parker *et al.*, 2004). This change in breeding objectives does not necessarily mean that behaviour is less essential for dog owners today than it was prior to modern breed creation, only that breeders in many dog breeds tend to place more relative importance on other types of traits.

The most important role for dogs in the western world today is as companion animals (Hubrecht, 1995). Dogs kept as family pets have to cope with various situations in their daily life. They are frequently exposed to noisy and crowded environments, and often have to interact with people, dogs and other animals. High levels of fear, anxiety or aggressiveness in dogs cause difficulties both for the dogs, from a welfare perspective, and for the dog owners, for example by inducing limitations in their everyday life. Aggressive behaviours can even be seen as a problem for society as a whole, and many countries have adopted far-reaching legislation aiming at limiting problems with "dangerous dogs" (Hundansvarsutredningen, 2003). According to McGreevy (2008), unwanted behaviour is the most common reason for euthanizing dogs in the developed world.

Behavioural traits in dogs also represent an economic value, for example when herding dogs are utilized by livestock farmers, or hunting dogs by hunters. Most of the studies in this thesis are on Swedish dog populations. Compared to almost all other countries, the proportion of Swedish dogs used for hunting or hunting trials is very high, about 25% (Egenvall *et al.*, 1999). Arnott *et al.* (2014) estimated the median value of the work performed by an Australian stock herding dog over the period of its working life to AU$40,000. In many parts of the world, police, customs and military authorities as well as guide dog schools report difficulties in finding dogs suitable for service (e.g., Goddard and Beilharz, 1982; MacIsaac *et al.*, 2005; Tjänstehundsutredningen,

2005; Vanderloo, 2005; Slabbert, 2008). In Sweden, there are more than 4500 working/service dogs, generating a substantial benefit for the society, for which training and/or use are partly or fully financed by the society (Tjänstehundsutredningen, 2005). The Swedish governmental inquiry Tjänstehundsutredningen (2005) described how authorities utilizing these dogs, such as police, military and customs, reported increasing problems of finding dogs with appropriate temperament.

In summary, dog behaviour is essential and heritable. It is also possible to breed for (Willis, 1995). Therefore, behaviour is important to consider when selecting breeding animals.

### 2.2.2 Dog breeding

The founder event of a breed, which typically happened less than 200 years ago, commonly involved only few dogs, and since then the only dogs generally allowed to be registered as members of a breed are those whose parents are registered (Ostrander and Wayne, 2005). Within these (many small) closed populations, popular sires have been widely used as stud dogs, and the breeding practice has often involved high selection intensities for breed-specific traits according to the breed standards and systematic inbreeding by mating of close relatives (McGreevy and Nicholas, 1999; Ostrander and Wayne, 2005). As a result, many dog breeds are characterized by limited genetic variation and a small effective population size, as indicated by the dog genome structure with long haplotype blocks within breeds (Lindblad-Toh *et al.*, 2005).

Defining a breeding goal is never an easy task, but compared with livestock breeding it is probably even more difficult when breeding dogs. One reason is that for production traits, the importance of a trait often can be measured in economic terms. Dog breeding is usually a hobby and not mainly driven by economic incentives, and the breeding goal relies on much more vague criteria. Traditionally, the main breeding goal for pedigree dogs seems to have been dog show success and thus primarily containing conformation and external characteristics (Willis, 1995; McGreevy and Nicholas, 1999; Mäki *et al.*, 2005; Liinamo and van Arendonk, 2006; Svartberg, 2006; McGreevy, 2008; Rooney, 2009). There are also populations where the focus is on functional traits, such as hunting or herding skills, or on traits relevant for, e.g., guide or police dog service. Whether breeding for appearance or functional traits or both, the breeding goal is often difficult to define, and selection is primarily based on subjective measures (if measures exist at all).

Compared with livestock breeding, dog breeding is decentralized in the way that many breeders relatively independently selects which dogs to use for

breeding. Each breeder usually has only a small production of puppies. There are a few examples where EBVs are being used to select for hunting traits in dogs (e.g., Finnish Hound in Finland (Liinamo, 2004) and Drever (Swedish Dachsbracke) in Sweden (Cederström *et al.*, 1994)), for hip and/or elbow dysplasia in some countries (Mäki, 2004; Stock and Distl, 2010; Swedish Kennel Club, 2014), for temperament in Collie in Sweden (Paper IV), and for traits regarded important for guide dog functionality, at a few guide dog schools in the U.S. (Jane Russenberger, 2013, pers. comm.). In general, however, EBVs are still rarely used by dog breeders. Instead, selection based on phenotypes is the most common method. This is especially unfortunate when breeding for traits with low heritabilities because the relative benefit of using BLUP can be expected to increase with decreased heritability by the use of information from relatives. Behavioural traits typically show low to medium heritabilities (Willis, 1995), whereas many traits of importance when breeding livestock show medium to high heritabilities, such as daily weight gain or back-fat thickness in pigs (Falconer and Mackay, 1996). Introduction of BLUP requires a certain degree of (infra)structure, for example a reliable pedigree and systematically recorded phenotypes.

## 2.3   Main issues

To breed systematically for behaviour in dogs, applying for example BLUP methodology, behavioural measurements are necessary. Behavioural data can be collected in many different ways, and the measurement method has implications for the breeding program. For example, measurements can be made in competitions or field trials (e.g., Karjalainen *et al.*, 1996; Hoffmann *et al.*, 2002; Correau & Langlois, 2005), when the dog is exposed to a standardized test battery (e.g., Wilsson & Sundgren, 1997; Ruefenacht *et al.*, 2002; Saetre *et al.*, 2006), or by observing the dog in everyday life, during training or while walking in different environments (e.g., Murphy, 1997; Schiefelbein, 2012, 2013). The judges can be extensively trained for the task (e.g., Ruefenacht *et al.*, 2002; Saetre *et al.*, 2006), persons regarded as skilful and competent but without formal training, or, if a dog owner questionnaire is used, the dog owners (e.g., Liinamo *et al.*, 2007). The measurements can be more or less objective, and the ratings can refer to behaviours displayed in a specific situation or to an overall interpretation indicating the degree of expression of pre-defined traits (Wilsson & Sinn, 2012). These factors, and many more, can be expected to affect the usefulness of the measurements when used in a breeding program. It is, however, not necessarily obvious in which direction the usefulness will be affected. For example, competitions may attract

dog owners to actually participate in the recording program. On the other hand, prospects of winning a prize probably makes the dog owners more prone to train their dogs to perform their best, thereby masking the genetic potential which was the aim to measure.

In Sweden, as well as in the rest of the world, methods for measuring different types of behaviours and behavioural traits are frequently used, for example, herding, hunting and working dog trials, temperament tests for puppies, for young dogs or for adults, questionnaires, etc. Many of these were originally designed to be used for selection of breeding animals, and some are still used for this purpose. Yet only a very small fraction of them have been analysed from a breeding perspective. This may cause breeders' confidence in the test to decrease over time. As a consequence, the measurement method might with time not be used for selection, implying that the value of testing dogs decreases rather dramatically regardless of how good the method is in itself. But the main problem is if measurements intended for selection of breeding animals do not show enough genetic variation, or if they are not genetically correlated to traits in the breeding goal. One consequence of inadequate measurements is unnecessary costs of recording phenotypes. Another effect, if the measurements are still used for selection, is that the room for selecting for other traits decreases without the intended genetic progress (or, in a worst case scenario, with an unfavourable genetic change). Therefore, three relevant questions to address are:

1. Do the measurements show genetic variation? If a temperament test or a field trial is supposed to function as a basis for selection, the measurements must show genetic variation. If they do not, it will not matter how advanced methods for genetic evaluation that are used, or how intense selection is; selection will not lead to the intended genetic progress.
2. Are the measured traits genetically correlated to the breeding goal traits? Sometimes, a measured trait or a selection trait is identical to a breeding goal trait, for example milk yield in cows. But sometimes the connection is less obvious. For example, in the Swedish temperament test Dog Mentality Assessment, the dogs are exposed to two slowly approaching persons dressed as ghosts (covered in white sheet and a hood over the head). Of course, the breeding goal has never been that the dogs should act in a certain way when exposed to ghosts. More likely, the breeding goal contains traits like aggressiveness or fearfulness, and the intention with the ghost subtest is to measure behaviours related to aggressiveness and fearfulness. Whether this is successful or not is impossible to know until the genetic correlations have been studied.

3. How should a method for measurement and genetic evaluation of behavioural traits be designed to function well for selection of breeding animals?

In this thesis, eight methods to measure behavioural characteristics in dogs were analysed. All but one were designed to be used for selection for breeding. For a majority of the studied methods, no genetic analysis has been previously published, and none of the methods have been studied based on the populations used in this thesis.

# 3   Aims of the thesis

The overall aim of this thesis was to investigate the prospects for improving dog behaviour by breeding. Genetic evaluations were therefore performed for a number of methods to measure different types of behavioural characteristics in dogs, thereby 1) investigating the potential for genetic progress for each method if used for selection, and 2) advancing our understanding of factors affecting the usefulness of behavioural measurements for breeding purposes. More specifically, the objectives were to:

➢ Estimate genetic parameters for behavioural traits based on: two consecutive versions of the Swedish Herding Trait Characterization; the English Setter field trials in Sweden and Norway; two types of ratings (behavioural ratings and subjective ratings) used in the Swedish Armed Forces temperament test for dogs; the Dog Mentality Assessment; and an extended version of the Canine Behavioral Assessment and Research Questionnaire.
➢ Study how the degree of objectivity and neutrality of a behavioural measurement affect the heritability.
➢ Study how different methods to define and compute composite traits based on the measured behavioural variables (items) affect heritabilities of the composite traits and/or their genetic correlations to traits in the breeding goal.
➢ Study the effect of across-country genetic evaluation on breeding value accuracies.
➢ Estimate genetic correlations between dog behaviour in a commonly used temperament test and in everyday life situations as perceived by the dog owners.

# 4 Summary of the studies

## 4.1 Behavioural data

In all papers behavioural data were analysed. In each paper, however, the data emanated from different measurement methods; the Herding Trait Characterisation (HTC) in Paper I, the English Setter field trial (ES FT) in Paper II, the Swedish Armed Forces temperament test (SAF TT) in Paper III, and the Dog Mentality Assessment (DMA) and an extended version of the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) in Paper IV. Also breeds differed between papers, as well as number of dogs with records. A summary of data is given in Table 1. All methods except the questionnaire had been designed mainly to function as tools for selection for breeding.

Pedigree data for the analysed breeds were obtained from the Swedish Kennel Club database and edited: duplicate ID numbers were removed, parents who did not occur as individuals were added as individuals with unknown parents, and dogs with the same name, mother and birth date were considered as the same individual and only one of the records was kept. Behavioural records belonging to dogs that did not exist in the pedigree were deleted, and impossible observations, such as score 6 on a scale from 1 to 5, were considered as missing.

Paper I was based on two consecutive versions of the Herding Trait Characterisation (HTC). The HTC − which is no longer in use − was a non-competitive method (no winner was nominated among the participating dogs) to describe how individual dogs typically expressed a number of traits considered important for herding ability. For example, the trait Balance described a dog's ability to work in balance with the handler, i.e. to take the position on the opposite side of the flock that affects the livestock to move towards the handler, and the trait Out-run how wide circle the dog made when

*Table 1*. Summary of behavioural data used in Papers I-IV

| Paper | Measurement method | Data collected by | Breed | No. of items recorded and analysed | Recording period (years) | No. of dogs with records |
|---|---|---|---|---|---|---|
| I | Herding Trait Characterisation (HTC), version 1 | Swedish Sheepdog Society | Border Collie | 17 | 1989-1995 | 1663 |
| I | Herding Trait Characterisation (HTC), version 2 | Swedish Sheepdog Society | Border Collie | 19 | 1996-2003 | 951 |
| II | English Setter field trial (ES FT) in Sweden | Swedish Setter Club for English Setter | English Setter | 6 | 2003-2010 | 685 |
| II | English Setter field trial (ES FT) in Norway | Norwegian English Setter Club | English Setter | 6 | 1994-2011 | 7175 |
| III | Swedish Armed Forces temperament test (SAF TT), behavioural ratings (BR) | Swedish Armed Forces | German Shepherd Dog | 25 | 2006-2012 | 873 |
| III | Swedish Armed Forces temperament test (SAF TT), subjective ratings (SR) | Swedish Armed Forces | German Shepherd Dog | 13 | 2006-2012 | 873 |
| IV | Dog Mentality Assessment (DMA) | Swedish Working Dog Association | Rough Collie | 33 | 1997-2010 | 2953 |
| IV | Questionnaire | The authors | Rough Collie | 95 | 2010 | 1738 |

moving from the handler to the balance point. In the first version, 17 traits were measured, 12 of which can be regarded as herding traits. The second version contained 19 traits, 14 of which being herding traits. Most traits in version 1 were also in version 2. The main difference between the versions was the structure of the score sheets used to record the traits; the predefined grading alternatives – in most cases 6-step scales – were almost always different between the two versions. In version 1, the intention was to measure the intensity or the strength of the expression of a trait. The descriptions used for the various grades were written as not to be interpreted as ''good'' or ''bad''. In version 2, the most desirable expression of a trait was placed in the middle of the scale. Another difference was that in the second version, the judges were given more freedom for their own interpretations, that is, the scales were more subjective.

In paper II, hunting traits measured in the English Setter field trial (ES FT) in Sweden and Norway were analysed. The English Setter is a pointing dog, a type of dog primarily used for hunting grouse; they search for the game, point and finally – on command from the hunter – makes it flush (fly), thereby giving the hunter a controlled opportunity to shoot. Examples of traits considered important were Hunting drive, describing the desire to find birds, Quartering (efficiency of search pattern), and Speed when searching. The study was based on six traits from each country (six from the Swedish trials plus six from the Norwegian). The trials were similar between countries, and the intention had been to define the six traits equally between countries and to assess them on equivalent 6-step scales indicating the degree of expression of each trait, and/or how desirable the degree of expression of the trait is compared to the ideal. There were, however, also some differences between countries. The Norwegian judges were more extensively trained, and in addition around half or the Norwegian trials were judged by two judges simultaneously, whereas Swedish trials were always only judged by one judge. Furthermore, for some of the six traits analysed, the Norwegian score sheet scales were considered slightly more objective and neutral.

Paper III was based on a temperament test developed by the Swedish Armed Forces. The Swedish Armed Forces temperament test (SAF TT) was in the form of a test battery containing 12 standardized subtests. In one subtest, for example, the test leader invited the dog to bite and pull a cotton rag, in another the dog was walked up and down steep metal stairs. The dog's behaviour during the test was simultaneously rated using two separate score sheets; in the first score sheet the rating method was termed "behavioural ratings" (BR), in the second "subjective ratings" (SR). The BR were based on the judge's observations of behaviours displayed in a specific subtest, and the

intention was to rate behaviour as objectively as possible; 25 BR were given using pre-defined 5-step scales containing typical behaviours characterizing each step of the scale for each item. The scales were arranged according to the intensity of the behavioural reactions. The SR were based on the judge's overall interpretation of how a dog behaved during the whole test (one of the ratings was based on only one subtest). When giving the 13 SR, the judge assessed pre-defined temperament traits, such as Courage (the absence of fearful behaviour toward real or imagined danger) or Curiosity (the tendency to explore and to investigate novel things) using 5-step scales (for one trait the scale contained 6 steps) indicating the degree of expression of each trait.

In Paper IV, two methods to measure behaviour were analysed: Dog Mentality Assessment (DMA) and an extended version of the Canine Behavioral Assessment and Research Questionnaire (C-BARQ). The DMA was a test battery of ten standardized subtests, during which the intensities of 33 behavioural reactions shown by the tested dog were rated according to a score sheet with 5-step scales for all 33 items. For example, in the subtest Play 1, three items described how the dog behaved when playing with a stranger with a rag: how interested the dog was, how the dog grabbed the rag, and how engaged the dog was in tug-of-war. All steps of the scales for all items contained short descriptions of typical behaviours. The intention when constructing the scales was to define each step of a scale as objectively as possible and to arrange the steps from low to high intensity of the behavioural reactions, i.e. a low rating corresponds to low intensity of the reaction. No judgment was made during the test whether a dog showed preferred behaviours or not.

The major part of the questionnaire was a Swedish translation of C-BARQ. The C-BARQ (Hsu & Serpell, 2003; Duffy & Serpell, 2012) contained 101 questions, 22 of which could be removed without potentially reducing reliability and/or validity; 21 of these were removed, leaving 80 items. To these original C-BARQ items, 15 questions regarding playfulness and sociability were added in accordance with Svartberg (2005). In the questionnaire, the dog owner was asked to rate their dog's typical behaviour in the recent past on a 5-step scale; either the frequency of certain behaviours (Never – Seldom – Sometimes – Usually – Always) or the intensity of the behaviour in defined situations (e.g., "No aggression: No visible signs of aggressive behaviours" to "Serious aggression: Snaps, bites or attempts to bite").

## 4.2 Statistical methods

Single-trait (Paper I) and both single- and multiple-trait (Papers II-IV) mixed linear animal models were used to estimate genetic parameters. Fixed environmental effects were tested for significance using SAS Proc GLM (Papers I-II) and Proc MIXED (Paper III) (SAS, 2008). To test random environmental effects for significance (Papers II-III), and to estimate (co)variance components (Papers I-IV) and breeding values (Paper II), the DMU software (Madsen and Jensen, 2010) was used. The final models for analysing measurements included the following effects in addition to random additive genetic effect of the individual and residual:

➢ HTC, both version 1 and 2 (Paper I): Fixed effects of sex and test year.
➢ ES FT in both Norway and Sweden (Paper II): Fixed effects of sex, type of trial, test year, test month and interaction between age and class of trial, and random effects of permanent environment and judge.
➢ SAF TT, both BR and SR (Paper III): Fixed effects of sex, training level, test age and test year − test location combination, and random effect of litter.
➢ DMA (Paper IV): Fixed effects of sex, year and month of test, and random effects of litter, judge, and test occasion. Age at test was included as linear and quadratic regressions.
➢ Questionnaire (Paper IV): Fixed effect of sex. Age when the questionnaire was answered was included as linear and quadratic regressions.

In Paper II, breeding values accuracies from single-trait within-country genetic evaluations were compared with accuracies for the same individuals from a bivariate across-country genetic evaluation.

In a previous study by Wilsson and Sinn (2012), principal component analysis was used on the SAF TT item phenotypes to define five and three composite traits − so-called underlying behavioural dimensions − from each rating method (BR and SR, respectively). In Paper III, these original behavioural dimensions were redefined by excluding items with heritabilities estimated at 0.00, and subsequently by excluding items that did not correlate well genetically to other items within the same dimension. In total, four dimensions were redefined by excluding one or two items from each. One of the original dimensions showed very little genetic variation and was excluded from all further analyses. After these modifications, the genetic correlations between the original dimensions and the redefined ones were estimated. Finally, genetic correlations among all seven (redefined) behavioural dimensions were estimated in a multiple-trait analysis.

In Paper IV, factor analysis, following Hair *et al.* (1998), was performed on the 33 DMA items using Proc Factor (METHOD=PRINIT) in SAS (2008), and five factors were extracted, representing five so-called personality traits. After orthogonal varimax rotation, the factor loading pattern indicated that 22 of the 33 items would be appropriate to use when computing five composite scores following the concept of summated scales (SS). A SS was calculated as an average of the standardized (mean 0, SD 1) values for the items judged to be good representatives for that factor (based on factor loadings). Furthermore, the loadings from the rotated solution were used for computing factor scores (FS). To compute a FS, all 33 items were included, weighed by their respective factor loadings for that factor. Thus, composite scores were constructed both as SS and as FS, and both types of scores were used in the further analyses. Based on previous studies, the questionnaire items were condensed into 18 so-called behavioural subscales scores, calculated as the average of the included items. Genetic correlations between questionnaire behavioural subscales and the two versions of the five DMA personality traits and the DMA item Gunshot avoidance were estimated in bivariate analyses.

## 4.3   Main findings

### 4.3.1   Genetic and environmental factors affecting behaviour

Average heritabilities for measured items within test varied from 0.10 to 0.32 (Table 2). Almost all studied measurement methods showed heritabilities or additive genetic variances significantly different from zero for a majority of the measured items. The only exception was the SAF TT BR, where only 44% of the measurements showed significant heritabilities. For the questionnaire no variance components were estimated for the items.

For four of the studied methods (SAF TT BR, SAF TT SR, DMA, questionnaire), the measured items were summarized into seven sets of composite scores (for three of the methods, two sets of composite scores were calculated for each). For five of these seven sets, all composite score heritabilities were significantly different from zero (Table 2). For both remaining sets, more than half of the composite score heritabilities were significant. Average heritabilities for composite traits within test varied from 0.16 to 0.20.

*Table 2.* Heritability estimates ($h^2$) and proportion of heritabilities significantly different from zero (Sign.) for measured items, and for composite traits computed based on subsets of items, for two versions of the Herding Trait Characterization (HTC); the English Setter field trials (ES FT) in Sweden and Norway; two types of ratings (behavioural ratings, BR, and subjective ratings, SR) used in the Swedish Armed Forces temperament test (SAF TT); the Dog Mentality Assessment (DMA); and an extended version of the Canine Behavioral Assessment and Research Questionnaire

| Measurement method | Recorded items | | | Composite traits | | |
|---|---|---|---|---|---|---|
| | N | $h^2$ [average (min, max)][a] | Sign. (%) | N | $h^2$ [average (min, max)][b] | Sign. (%) |
| HTC, version 1 | 17 | 0.32 (0.14-0.50) | 100 | ND | ND | ND |
| HTC, version 2 | 19 | 0.20 (0.03-0.41) | 84 | ND | ND | ND |
| ES FT, Sweden | 6 | Within: 0.10 (0.07-0.13) | 100[c] | ND | ND | ND |
| | | Across: 0.11 (0.08-0.13) | 100[c] | | | |
| ES FT, Norway | 6 | Within: 0.15 (0.08-0.18) | 100[c] | ND | ND | ND |
| | | Across: 0.15 (0.08-0.18) | 100[c] | | | |
| SAF TT, BR | 25 | 0.11 (0.00-0.21) | 44 | 4 | Original: 0.16 (0.08-0.22) | 75 |
| | | | | | Redefined: 0.18 (0.15-0.23) | 100 |
| SAF TT, SR | 13 | 0.14 (0.00-0.21) | 54 | 3 | Original: 0.18 (0.13-0.28) | 67 |
| | | | | | Redefined: 0.20 (0.12-0.28) | 100 |
| DMA | 33 | 0.14 (0.03-0.30) | 94 | 5 | SS: 0.19 (0.14-0.25) | 100 |
| | | | | | FS: 0.17 (0.13-0.21) | 100 |
| Questionnaire | 95 | ND | ND | 18 | 0.17 (0.06-0.36) | 100 |

ND = No data.

[a] Within: Univariate genetic analysis of national data alone; Across: Bivariate genetic analysis of joint Swedish and Norwegian data.

[b] Original: Composite traits defined according to Wilsson and Sinn (2012); Redefined: Composite traits redefined by exclusion of items with heritabilities estimated to 0.00, and subsequently by exclusion of items that did not correlate well genetically to other items within the same trait. SS: Composite trait scores calculated as summated scales; FS: Composite trait scores calculated as factor scores.

[c] Standard error for $h^2$ not estimated, but all additive genetic effects were significantly different from zero.

For all studied measurement methods, a majority of the items were shown to be under significant influence by at least one systematic environmental effect. In addition to fixed effects and the random genetic effects of the individual and residual, the models for analysing ES FT, SAF TT and DMA included one or several systematic random environmental effects. In Figure 2, the relative influence of these random effects is illustrated as averages over the measured items.
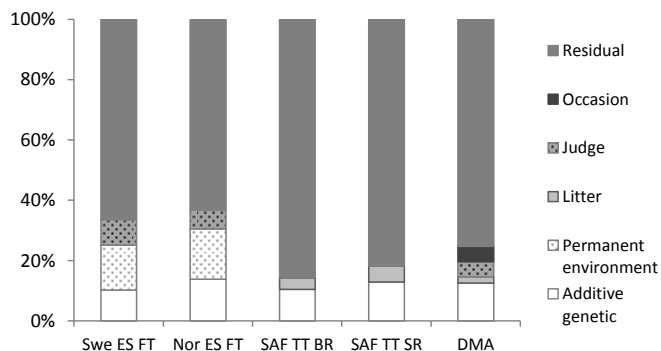


*Figure 2*. Relative influence of random effects (averages over the measured items) in the Swedish (Swe) and Norwegian (Nor) English Setter field trials (FT), the Swedish Armed Forces temperament test (SAF TT), both for behavioural ratings (BR) and subjective ratings (SR), and in the Dog Mentality Assessment (DMA).

### 4.3.2 Genetic correlations between temperament test result and everyday life behaviour

In Paper IV it was shown that each of the five DMA personality traits (computed as SS), as well as the DMA item Gunshot avoidance, were significantly genetically correlated with at least two of the 18 questionnaire subscales. The strongest genetic correlations for the DMA personality traits were for DMA Curiosity/Fearlessness with questionnaire Non-social fear (-0.70), DMA Playfulness with questionnaire Human-directed play interest (0.63), DMA Chase-proneness with questionnaire Chasing (0.73), and DMA Sociability with questionnaire Stranger-directed interest (0.87). The DMA personality trait Aggressiveness was not significantly correlated with any of the questionnaire subscales measuring different aspects of aggressiveness in everyday life. The correlation between the DMA item Gunshot avoidance and the questionnaire subscale Non-social fear was estimated at 1.00.

### 4.3.3 Score sheet structure

In Paper I, the heritability estimates of the traits measured in HTC version 1 were substantially higher than those of the corresponding traits in version 2 (Table 2). If selecting on phenotype, all else being equal, the potential genetic

progress would be on average 50% higher if using version 1 over version 2. In version 1 of the HTC the scales in the score sheets were considered more clear, objective and neutral. Similarly, in Paper II, the heritability estimates of traits measured using the slightly more objective and neutral scales in the score sheets used during the Norwegian ES FT were higher than the corresponding Swedish ones (Table 2). In Paper III, each BR behavioural dimension showed a high genetic correlation (0.89-0.98) with at least one of the SR dimensions. Pairwise comparisons of heritabilities between the more objectively measured BR dimensions and the corresponding SR dimensions showed that neither method performed systematically better than the other from a heritability perspective. Among totally four comparisons, the BR method had higher heritability estimates for two and the SR method for two.

### 4.3.4 Summarizing correlated measurements into composite scores

In Papers III and IV, the heritability estimates in general were higher for the composite traits than for the items used for calculating them. The differences between composite trait heritabilities on the one hand, and average heritabilities for the items used for calculating composite trait scores on the other, were not significant. The composite trait heritability estimates were however higher for all twelve pairwise comparisons (Figure 3), indicating that summarizing correlated measurements into composite scores indeed has a positive effect on the heritabilities.
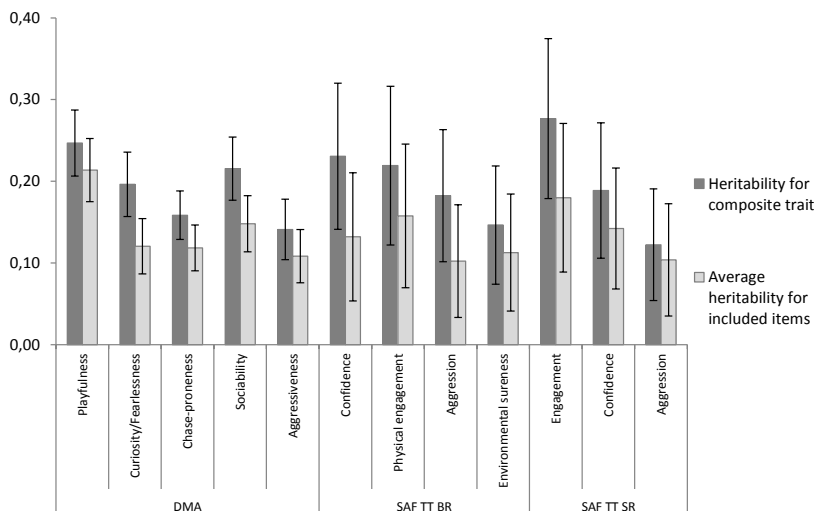


*Figure 3.* Heritabilities and standard errors for composite trait scores, and averages of heritabilities and their standard errors for the items used for calculating each score, from the Dog Mentality Assessment (DMA), and from the Swedish Armed Forces Temperament Test (SAF TT), both the behavioural rating (BR) and the subjective rating (SR) score sheet.

### 4.3.5 Different strategies to define and compute composite trait scores

In Paper III, the four redefined behavioural dimensions remained genetically almost identical to the original ones: the genetic correlations varied between 0.95 and 1.00. For SR aggression the heritabilities were estimated at 0.13 and 0.12 for the original and the redefined versions, respectively. For all remaining comparisons, the heritability estimates were higher for the redefined dimensions (0.20 and 0.23, 0.08 and 0.15, 0.13 and 0.19).

Regarding DMA personality traits computed as SS and FS in Paper IV, there were strong correlations (0.85-1.00) for all random effects (genetic, litter, judge, occasion and residual) between corresponding SS and FS (SS Playfulness versus FS Playfulness, SS Curiosity/Fearlessness versus FS Curiosity/Fearlessness, etc.), indicating that they can be considered as more or less the same traits. For all five pairwise comparisons of heritability estimates, the SS method resulted in equal or greater estimates compared with the FS method (with the exception for Aggressiveness: 0.14 and 0.15), mainly due to greater residual variance for the FS (on average 36% higher). The genetic correlations between both versions of the DMA traits (SS and FS) and everyday life behaviour of the dogs as described by the owners in the questionnaire were similar, and neither method to calculate underlying DMA traits succeeded systematically better than the other in this respect.

### 4.3.6 Joint genetic evaluation

The calculations of breeding value accuracies in Paper II showed that especially Swedish breeders would benefit substantially in terms of accuracy of breeding values from utilizing across-country data: for all traits in both countries, the average accuracy increased when the breeding values were predicted using a joint evaluation (bivariate analysis) instead of a univariate analysis on national data alone. For dogs with Swedish trial results the average increase was 19% (for dogs with Norwegian trial results the average increase was only 1%). Also minimum and maximum breeding value accuracies increased for all traits in both countries when data were merged.

# 5 General discussion

## 5.1 The analysed behavioural measurement methods can be used in genetic evaluations

### 5.1.1 Increased genetic progress

Taken together, the studies in Papers I-IV show that behavioural traits – herding as well as hunting and general temperament traits – are influenced by genetic factors. The results also showed that it is possible to achieve genetic progress by utilizing the studied measurement methods for selection of breeding animals. For all studied methods, a majority of the items were influenced by at least one systematic environmental effect, which therefore should be taken into account. Selection on the individuals' phenotypic records is the most common method in dog breeding today. Using a BLUP animal model to estimate breeding values would potentially increase the annual genetic progress by adjusting for systematic environmental effect as well as taking information from all tested relatives' performance into account. Compared with the current situation, it would thus be possible to use the studied measurement methods in a genetic evaluation to achieve a faster improvement in the genetic level for herding and hunting traits among Border Collies and English Setters, respectively, and for temperament traits in the German Shepherd Dog and the Rough Collie.

In a simulation study, selection on BLUP breeding values for hip dysplasia in dogs resulted in a substantially faster genetic progress compared to when selection was on phenotypic records (Malm *et al.*, 2013). Behavioural traits typically show lower heritabilities than hip dysplasia (for example 0.37-0.42 for hip dysplasia in Rottweiler and Bernese Mountain Dog in Finland and Sweden (Mäki et al., 2002; Malm et al., 2008) and 0.10-0.32 for behavioural traits (averages within measurement methods studied in Papers I-IV)), and the lower the heritability, the greater the expected relative benefit of using BLUP.

The average improvement in genetic progress for the hunting traits studied in Paper II was calculated to be 66% in Sweden, and 87% in Norway, if using BLUP breeding values over phenotypes for selection. Even though this might seem like rather dramatic differences, they rely on the assumption that when using phenotypes for selection, the records have been adjusted for the fixed effects of sex, test year, test month, type of trial, and interaction between age and class of trial. Because no such adjustment is likely to be made in case of phenotypic selection, the real difference is probably substantially greater than 66% and 87%.

### 5.1.2 Comparison with previous studies

The results from Papers I-IV are well in agreement with those from previous studies in that behavioural traits are heritable but typically show low-to-moderate heritabilities. The heritabilities for herding traits found in Paper I were however high compared with the few other genetic studies available on sheepdogs and herding traits. Based on 2745 results of 337 Border Collies, Hoffman *et al.* (2002) estimated heritabilities for various herding traits from close to zero to 0.13. Swenson (1983) used data from the predecessor of HTC and estimated heritabilities for herding traits, all below 0.20.

Brenøe *et al.* (2002) studied seven traits recorded during field trials for three pointing dog breeds, German Shorthaired Pointer, German Wirehaired Pointer and Brittany Spaniel (Breton). They estimated heritabilities for the seven traits at 0.09-0.28, which is similar to the estimates from Paper II. Also Vangen & Klemetsdal (1988) obtained similar heritabilities (0.09-0.22) for four traits defined and recorded in a similar way as in the study by Brenøe *et al.* (2002), but measured in English Setter.

Heritabilities for temperament traits, measured using a test battery similar to the SAF temperament test (Paper III) or DMA (Paper IV), have been published in a handful of studies, and are generally well in concordance with the results in Papers III and IV. For traits defined and rated similarly as the SR items in the SAF test, heritabilities have typically been estimated at 0.10-0.30 (Wilsson and Sundgren, 1997; Ruefenacht *et al.*, 2002; van der Waaij *et al.*, 2008; Meyer *et al.*, 2012). Liimatainen *et al.* (2008) presented somewhat lower heritability estimates (0.04-0.13), in a study based on 2327 Rottweilers tested in an official behaviour test in Finland. Saetre *et al.* (2006) analysed DMA results from German Shepherd Dogs and Rottweilers. Their heritability estimates for the items were quite similar for the two breeds and varied between 0.04 and 0.19. Strandberg *et al.* (2005) estimated heritabilities for four of the five DMA personality traits at 0.09-0.26.

There are very few studies in which genetic parameters have been estimated for the C-BARQ items or subscales analysed in Paper IV. Liinamo *et al.* (2007) presented highly varying heritability estimates, some of them extremely high, for different C-BARQ scores related to aggressiveness in Golden Retriever dogs. However, their analyses included relatively few individuals (N=115-316), which in addition were pre-selected; the subjects had been recruited to the study either because they had shown aggressive behaviour, or because they were closely related to an aggressive dog. Several of the heritability estimates were 0.00 or 1.00, and for roughly half of the analyses no standard error could be obtained. The authors emphasize that the results should be approached with caution, and that the conclusions that can be drawn from the study are limited. In a master thesis study, Schiefelbein (2012, 2013) collected C-BARQ data on Labrador Retrievers, Golden Retrievers and German Shepherd Dogs that were six or twelve months old. The dogs had been bred at two American Guide dog schools. Heritabilities for the subscales were estimated at 0.00-0.47. Only every seventh estimate was > 0.10, and thus the heritabilities were in general lower compared with the results in Paper IV.

### 5.1.3 Correlations between measured traits and breeding goal traits

The more accurate selection, the faster genetic change can be expected. Implicitly, genetic change in a selection trait is favourable, but if this trait is not genetically correlated to the breeding goal, no genetic change for the breeding goal will take place. Thus, for a temperament test to be useful for selection, the measurements have to be genetically correlated to traits in the breeding goal. In a worst case scenario the selection trait is unfavourable correlated to a breeding goal trait, which – if not considered – could result in genetic change in an undesirable direction. For example, Mackenzie *et al.* (1985) found indications of an unfavourable genetic correlation between temperament and hip dysplasia in German Shepherd Dogs bred and evaluated by the United States Army's Division of Bio-Sensor Research; a desirable temperament score was negatively correlated with a desirable hip dysplasia score; dogs with a desirable temperament score tended to have a poor hip dysplasia score and vice versa.

In Paper IV it was shown not only that DMA can be used to achieve genetic change for the DMA personality traits. In addition, selection based on the DMA traits would bring about a genetic change for what was considered breeding goal traits, measured in the dog owner questionnaire. Fear-related problems are common among Rough Collies in Sweden. This is a problem not only for the dogs, from an animal welfare perspective, but also for the owners by inflicting limitations in their everyday life. Therefore, the questionnaire

subscale Non-social fear was considered as the most important trait in the breeding goal. The high and significant genetic correlations between the questionnaire subscale Non-social fear and the DMA trait Curiosity/Fearlessness (-0.70, SE 0.10) and the DMA item Gunshot avoidance (1.00, SE 0.12) show that the temperament test DMA could be an effective tool for selection of breeding animals with the goal to decrease everyday life fearfulness in the Swedish Rough Collie population. DMA can also be used for breeding for other everyday life behavioural traits, such as Human-directed play interest, Chasing, Stranger-directed fear and Separation-related behaviour.

Heritabilities for the questionnaire subscales were similar to those of the DMA personality traits. For the questionnaire subscale Non-social fear the heritability estimate (0.36) was even higher than for any of the DMA personality traits. A justified question is if it would not be better to select directly on the highly heritable breeding goal trait Non-social fear rather than on correlated DMA traits. If test results did not exist (which indeed is the case for most dog populations in the world), a routine genetic evaluation based on dog owner questionnaire results could be considered. In the Rough Collie case, however, where a high proportion of the dogs are tested in the DMA, selection based on DMA test results is recommended. A risk of using a questionnaire as a basis for routine genetic evaluation is that the reliability of the answers with time will become compromised. Basically, it is likely easier and more tempting for breeders to manipulate the breeding values of their dogs by convincing their puppy buyers to give certain answers in the questionnaire, than to bring about improved behavioural reactions in a standardized test like the DMA.

## 5.2   Recording methods

One of the aims of the thesis was to compare some measurement characteristics from a heritability perspective. When measuring a certain trait or behavioural response, the measurement error is influenced by how the measurement is conducted. Thus, the heritability can differ between measurement methods, even if referring to the same behaviour or trait. It could for example be hypothesized that the more objective a measurement, the higher the heritability. The objectivity of a measurement here refers to the rating alternatives in the score sheet scales. For example, in both versions of the HTC, the herding trait Effective working distance was measured using a 6-step scale. Effective working distance was defined as the distance between dog and livestock where the livestock became affected by the dog and started to move away. In HTC version 1, the distance was given in meters (0-1; 1-2; 2-3; 3-5; 5-10; >10). In version 2, the six rating alternatives in the scale were "Fails to

move the animals regardless of distance", "Needs to be very close"; "Needs to be relatively close"; "Needs a medium distance"; "Can move animals from a long distance"; "Can move animals from a very long distance". The former scale leaves less room for interpretation – it is more objective – and should therefore generate higher heritability for the trait Effective working distance. On the other hand, the situation is more complex in the way that the working distance is affected not only by the dog, but also by the livestock. Thus, the latter way of measuring might benefit from allowing for the judge to rate the dog given the behaviour of the livestock. Vazire *et al.* (2007) argued in favour of the supposedly more subjective "Trait ratings" over "Behaviour codings" when measuring personality in animals, partly because Behaviour codings "may reflect other characteristics of the environment (e.g., situational influences), not personality".

The results from Paper I and, to some extent, Paper II, indicate that the heritability tends to increase with the objectivity of a measurement, while the results from Paper III are more ambiguous. In Paper I, a major reason for the higher heritabilities in version 1 of the HTC was assumed to be due to the differences in how the score sheets were designed; in version 1, definitions of classes were more clear, objective and neutral. In Paper II, one explanation for the higher heritability estimates in Norwegian compared with Swedish ES FT could be the slightly more objective scales for some traits in the Norwegian score sheet. There are, however, alternative explanations. First, the estimates are not from the same population, and the difference can be due to higher genetic variance in the Norwegian population. Second, the Norwegian judges are more extensively trained than the Swedish counterparts, and in addition around half or the Norwegian trials are judged by 2 judges simultaneously, whereas Swedish trials are always judged by one judge only.

It can also be hypothesized that the heritability is affected by how neutral a measurement is (i.e., whether a dog is rated without the judge passing value judgments, or in terms of showing wanted or unwanted behavioural characteristics), standardization of testing routine and training of the involved personnel. For example, Vazire *et al.* (2007) concluded that a measurement can probably be made more reliable by training observers extensively and by providing specific definitions of behaviours and traits being measured. Besides the fact that rating dogs in terms of "good" or "bad" is not in accordance with this conclusion (to provide specific definitions of behaviours and traits), also other mechanisms may reduce the heritability if a measurement lacks neutrality. It is probably more difficult to remain objective if the score sheet forces you to evaluate and tell the owner how good or bad a dog is, rather than just in a neutral manner describe its temperament traits or how prone it is to

express different behaviours; judges might be reluctant to give dogs the "worst" grades. In Paper I, this was considered an important reason for the more extensive use of the whole score sheet scales in HTC version 1 compared with version 2. Also, the results in Paper II indicated that the judges tended to regress their assessments towards what was considered desirable. This might have two types of negative consequences. First, the full phenotypic variation will not be captured. Second, judges might differ in how influenced they are by circumstances other than how the dogs actually behave, and this type of judge variation cannot be easily adjusted for in a genetic evaluation.

For all studied measurement methods (Papers I-IV), there are examples of items showing comparably low phenotypic variation. In some cases, this might partly be a result of non-neutral score sheet scales according to the reasoning in the previous section. In other cases, the low variation is probably because the rating scale was not well adapted for the population in which it was used. If the phenotypic variation is not captured well, the likelihood of revealing genetic variation decreases. In Paper III, there are indications that (some of) the SAF behavioural measurements can be carried out in a better way by re-defining the scales used for rating the dogs' behaviours, for example by merging classes that are rarely utilized and by splitting classes to which a high proportion of the dogs are rated.

In summary, no simple and straightforward conclusions have been reached. On the other hand, those results pointing in a certain direction (primarily in Paper I), indicate that a more objective and neutral score sheet indeed is to prefer from a heritability perspective, and no results seem to indicate the opposite.

## 5.3   Summarizing measured items into composite traits

### 5.3.1  Why composite traits?

The average of several repeated measurements of a trait can be expected to show a smaller measurement error than a single measurement of the same trait. In Papers III and IV, the measured items were summarized into composite traits. The measurements were, however, not repeated measurements of the same traits. Instead, multivariate methods (principal component analysis and factor analysis) were used to define underlying components or factors, to which a number of items were correlated. Based on how strongly the items correlated to a factor, they were used to compute scores for the underlying traits. In one way, summarizing items into a composite score based on factor analysis is similar to averaging repeated measurements; the items that correlate strongly to a factor are likely to be correlated also to each other, and the

composite score is then not that different from the average of repeated measurements of the same trait. As expected, the composite traits showed higher heritability estimates than the items used for calculating them, and the reason is likely decreased measurement error due to repeated measurements.

There might be a similar explanation as to why the heritabilities of the HTC in Paper I, especially for version 1, are higher than in most other studies where heritability estimates of dog behaviour have been presented. Because the measurement for each trait is the result of repeated observations over eight to ten occasions, the rating can be regarded as an average of several repeated measurements. Similarly, the questionnaire heritabilities (Paper IV) probably benefitted from the fact that the dog owners had the opportunity of observing their dogs over a long period of time.

An advantage of using factor analysis to define fewer underlying traits, and then computing scores for the traits and basing selection on these, is that it is a convenient way to reduce the number of selection traits, thereby making selection more comprehensible. Another benefit of using several different measurements to define and compute an underlying trait, is that they may capture different aspects of the trait; they might be measured under different conditions (for example in the SAF TT where items from four different subtests were merged into the underlying trait Confidence) or by using different scales referring to different types of behaviours (for example in the DMA when merging startle reactions and exploratory behaviour into the underlying trait Curiosity/Fearlessness). Compared to repeated measurements of the same trait, this should improve the prospects to breed for traits that are stable over time and across similar situations, rather than for very specific behavioural responses valid only under certain conditions.

### 5.3.2 Two different concepts of computing composite trait scores

In Paper IV it was shown that the method used when computing scores for underlying traits – SS or FS – might influence the heritabilities of the traits. The SS method to compute DMA personality trait scores seemed to perform at least as good the FS method; estimates of heritabilities and genetic correlations between DMA results and everyday life behaviour as described by dog owners were generally equal or greater for the SS. Because they were also considered easier to compute and to explain, SS are the first choice in a breeding program for Rough Collie based on DMA data.

The FS showed greater residual variance than the SS. On the one hand, inclusion of all 33 original DMA items to calculate all 5 FS could have been expected to reduce residual variance with greater heritabilities as a result (when calculating SS, only 3 to 7 items were used to calculate each SS). On the other

hand, many items are only weakly correlated to each other and inclusion of all items when calculating FS apparently increased the residual variances and thus had a negative influence on the FS heritabilities.

### 5.3.3 Composite trait definition based on phenotype or on genotype?

When Wilsson and Sinn (2012) defined five behavioural dimensions traits based on the BR and three based on the SR, the purpose was to predict training success based on the composite trait scores and environmental factors. To predict the future success of a given dog, it makes sense to use a principal component analysis based on the phenotypic correlations among ratings. However, the genetic correlation between two traits can differ both in size and in sign compared with the corresponding phenotypic correlation (Falconer and Mackay, 1996). Consequently, it is not self-evident that a principal component analysis based on phenotypic records is optimal when constructing composite traits, if these are to be used for selection of breeding animals. In Paper III, one reason to why the heritability estimates in general became higher when the composite traits were re-defined based on genetic parameters, is likely different correlation structure between items on phenotypic and genotypic level. Another reason is the removal of non-heritable items. In conclusion, aggregating behavioural variables based on phenotypic correlations may be suboptimal when defining dimensions for breeding purposes; taking genetic parameters into consideration may lead to higher heritabilities for the aggregated traits.

## 5.4 Genetic evaluation

### 5.4.1 Systematic environmental effects

The results from Papers I-IV showed that a dog's sex and age affects its behavioural traits. This has previously been demonstrated in many studies (e.g., Karjalainen *et al.*, 1996; Strandberg *et al.*, 2005; van der Waaij *et al.*, 2008). The results also indicate that if enough dogs per litter have been tested, litter should be included as random effect to account for that litter mates are exposed to the same environment.

   Test month had a significant effect for a majority of the ES FT hunting traits (Paper II), and also, in agreement with Strandberg *et al.* (2005), for most DMA personality traits (Paper IV). Interestingly, test month was significant for only one of seven composite traits in the SAF TT (Paper III), a test very similar to the DMA. One major difference is however that DMA is performed outdoors, whereas SAF TT takes place mainly indoors. Van der Waaij *et al.* (2008) found a significant effect of test season when studying another test

similar to DMA and SAF TT, namely the temperament test previously used by the governmental Swedish Dog Training Centre (the centre does not exist anymore). They hypothesized that season may have influenced the test results due to seasonal fluctuations in serotonin and dopamine concentrations. The presence of a season effect in the outdoor tests, and the absence of such an effect in the indoor test, indicates that month or season of test affects behaviour more directly; the dogs tend to show different behavioural responses depending on, for example, temperature, whether or not the trees have leaves, or if there is snow on the ground or not, or some other factor in the environment that is present at the same moment as the measurement is made.

The effects of judge and year of testing are in agreement with previous studies where these effects have been tested (e.g., Strandberg *et al.*, 2005; Meyer *et al.*, 2012). The fact that calendar year significantly affects the dogs' results indicates that there are variations over time in how the measurements are made. There may for example be differences in how the actual testing is conducted, or changes in score sheets or definitions of traits. As long as these variations only affect the level of a rating, it can be adjusted for by inclusion of the effect of test year in the statistical model used for the genetic evaluation. If the variations also indicate that the measurements actually refer to different traits/behaviours depending on test year, it may become difficult to analyse measurements from different time periods in a univariate model. Similarly, the effect of judge should be included in most models for genetic evaluation of behaviour in dogs. More extensive education of judges to increase inter- and intra-rater reliability might be called for, but this has not been studied within the scope of this thesis. Using a BLUP model including the effect of judge will show how each judge judges relative the others. This will make objective feed-back to the judges possible, and also indicate if more education is required.

In the Norwegian ES FT, about half of the trials were judged by two judges making a joint evaluation. In Paper II this was regarded as a reason for the lower judge and error variances for Norwegian measurements, resulting in higher heritabilities. Maybe the Norwegian system would become even better if allowing each of the two judges to make an independent assessment, rather than the two of them making one joint assessment. A benefit of making separate assessments is that it becomes easier to correct for judge in the mixed model, because the number of levels for the factor judge then will become equal to the actual number of people judging rather than the number of unique judge combinations.

### 5.4.2 Unexplained environmental variation

Even if the heritabilities found in Papers I-IV are similar to those in previous and comparable studies, the unexplained variation – due to for example measurement error and unknown lifetime history events of an individual – is substantial in relation to the additive genetic variation. There are several possible actions to reduce the random error variation. Some of these actions refer to how the actual measurement of the phenotype is done and how the measured items are condensed into composite traits, and these aspects were discussed in 5.2 and 5.3. It would also be beneficial from a breeding perspective if more environmental factors influencing the test result could be identified and registered. This would make it possible to reduce the residual variance by including these new effects in the model, thereby generating a potentially faster genetic progress. Examples of factors that often seem to be neglected but that could be registered and tested for significance are weather conditions during test (if outdoors), number of spectators, personnel involved besides the judge, and geographic location. Preliminary analyses of a new Swedish temperament test for dogs, the ABC test (Assessment of Behaviour in Canines), indicated that rainfall, wind, thunder and temperature affected how the dogs behaved in standardized test situations (Svartberg, 2013).

The dog owner can be expected to have quite an influence on a dog's behavioural traits. Unfortunately it is usually not possible to adjust for owner in the model, because most owners – at least in the data analysed in Papers I-IV – are represented with only one dog, making it impossible to separate the owner effect from the residual. One way to at least partly get around this problem would be if each owner could be attributed one or several owner characteristics expected to influence the measured traits. Examples of such characteristics could be skilfulness, previous experience with dogs, sex, age, personality, ambition level concerning his/her dog ownership, relation between dog and owner, etc. Some of these owner characteristics would be easy enough to register (sex, age…), others more complicated (for example ambition level or skilfulness).

When testing dogs in the SAF TT, the dog's training level was registered. Training level was an estimate made by the judge of how much an individual dog had been trained by its puppy raiser. The estimate was based primarily on the number of training sessions arranged by SAF that the dog had participated in prior to test. In total, four sessions were arranged. Thus, training level can be seen as a characterization of how ambitious the puppy raiser has been. It might seem like a blunt way of characterizing an owner effect, but the fixed effect of training level was significant for all behavioural dimensions ($P<0.01$). A model with only fixed effects (sex, training level, test age and test year–test location

combination) explained on average 17% of the variation in the behavioural dimensions. If training level was excluded, the model explained on average only 10%, indicating that training level is an important factor. Similarly, Lindberg *et al.* (2004) showed that a dog's training level significantly affected hunting behaviour in Flatcoated Retriever.

Viklund (2010) discussed statistical models for estimating genetic parameters for competition traits in horses. There is a dependency between quality of the horses and quality of the rider in the way that the best riders tend to ride the best horses. If not including rider as a fixed effect, the genetic variance will therefore likely be biased upwards, and downwards if included. A similar dilemma with gene by environment covariance is probably present for at least some of the analysed traits in Papers I-IV, potentially complicating the prospects of including owner characteristics as fixed effects in the model. For example, Svartberg (2002) showed that some of the DMA personality traits are related to the owner's success in working dog trials with previous dogs. It is therefore not unlikely that people interested in working dog training tend to obtain dogs who are genetically predisposed to show high degrees of these traits. If these people at the same time are more prone and capable to develop these traits than less interested dog owners are, this would mean that the genetically "best" dogs in general were made to look even better at test. As a result, if this owner effect is not adjusted for, the genetic variance will be overestimated. And, vice versa, if the owner effect is adjusted for, the risk is that the genetic variance is underestimated.

### 5.4.3 Cooperation between countries

A major problem in dog breeding is that population sizes often are small (Lindblad-Toh *et al.*, 2005). As long as different breeds are treated as separate populations – even when they are morphologically and behaviourally very similar – the consequence is that for many breeds it is challenging to avoid high inbreeding rates. This also means that the room for selection is limited. In addition, if the aim is to breed systematically, the dogs available for selection are restricted to the ones that have their phenotypes recorded. This group is usually substantially smaller than the total number of dogs registered, limiting the room for selection even more. For example, compared to the situation in many other dog breeds, the number of Rough Collies subjected to the DMA is exceptionally large in relation to the number of registered dogs – between 25% and 50% of the dogs registered in the Swedish Kennel Club each year. Still, more than half of the dogs born do not end up among the selection candidates regarding behaviour. Therefore, two very important actions to strive for are to increase population sizes and to phenotype a higher proportion of animals

within populations. One obvious solution to the former problem would be to simply allow matings across similar breeds (which, for political reasons, might turn out not to be so simple). Another possibility is to capitalise on the fact that there often are genetic connections between a breed in one country/kennel organisation and the same breed in another country/organisation, due to use of the same or related dogs in breeding. As in Paper II, this can be exploited by merging the pedigrees. A prerequisite for this to be possible is that the animals have or can be assigned correct and unique id numbers. This would open up the possibility of making an across-country genetic evaluation, utilizing pedigree and phenotypic information from both populations. In comparison, international genetic evaluations based on an animal model have been carried out for the Swedish and Norwegian cold-blooded trotter since 1994 (Olsen *et al.*, 2012), and for dairy cattle in Denmark, Finland and Sweden since 2006 (Pösö *et al.*, 2006).

By connecting populations via a joint genetic evaluation, the number of selection candidates increases for all participating populations. In this way, cooperation between countries can be expected to make it possible to increase selection intensity and/or decrease the inbreeding rate. Cooperation between populations addresses also the fact that the accuracy of estimated breeding values can be expected to be influenced by number of individuals with records and pedigree completeness. As was shown in Paper II, an across-country genetic evaluation may increase the EBV accuracies rather dramatically. In other words, it is of great importance not only that as many dogs as possible within a population have their phenotypes recorded. It is also advised to investigate if related populations exist, ideally (but not necessary) for which similar behavioural traits are registered.

### 5.4.4 Alternatives to selection on BLUP breeding values?

A potential problem with BLUP is that closely related animals tend to get similar EBVs. This in turn means that the top animals may very well be closely related. If not carefully monitoring inbreeding rate and which animals that are used for breeding, and having agreements on population level on how high an inbreeding rate that can be tolerated, selection on EBVs might lead to problems related to a high inbreeding rate. A systematic way to address this dilemma would be to use optimum contribution selection, where the response of selection is maximized given a predefined restriction on the rate of inbreeding (Meuwissen, 1997). A limitation with optimum contribution selection is that a rather strong control over the breeding population is needed for it to work optimally. In practice, optimum contribution selection is therefore not likely easily implemented in dog breeding, because selection to a large extent is

controlled by individual breeders. It is not realistic to expect all hobby breeders in a dog breed to follow recommendations not only on which animals that should be used for breeding, but also on how many litters each dog should produce. The Swedish Armed Forces' breeding program is an exception, and in their case optimum contribution selection should be considered.

In livestock breeding, genomic selection is increasingly being used instead of BLUP for estimating breeding values. The costs for developing a genetic evaluation based on genomic information are however considerably higher compared to using BLUP. Genomic selection requires that the marker effects have been estimated in a reference population. The required size of the reference population depends on several factors, for example the heritability, effective population size and desired accuracy of selection (Goddard, 2009). Because dog breeding is characterized by small populations and limited economic resources, BLUP is currently a more rational and realistic choice in most cases; it is difficult to identify behavioural traits for which genomic selection would be superior enough to motivate the higher costs. Exceptions may exist, though, especially if considering other types of traits. For example, the serious heart disease myxomatous mitral valve disease is common among Cavalier King Charles Spaniel (Häggström *et al.*, 1992), but onset of the disease occurs late in life, usually after a dog has already been used for breeding. In this case the additional costs for genomic selection may be justified.

# 6 Conclusions

A majority of the behavioural measurements from the Herding Trait Characterization, the Swedish and Norwegian hunting trials for English Setter, the Swedish Armed Forces temperament test, the Dog Mentality Assessment, and the extended version of the Canine Behavioral Assessment and Research Questionnaire showed genetic variation and can be used for selection of breeding animals to achieve genetic progress. In most cases, systematic environmental effects have a significant influence on the behaviour. In combination with the low to moderate heritabilities, this suggests that selection of breeding animals based on individual performance is not ideal. Using BLUP breeding values would increase accuracy of selection and the potential genetic progress and is therefore recommended. To improve accuracy further, more environmental factors potentially influencing the behavioural measurements – for example weather conditions and dog owner characteristics – should be registered and, if appropriate, included in the model.

When correlated items from the Swedish Armed Forces temperament test and the Dog Mentality Assessment were used for computing composite traits, the heritability estimates were higher for the composite traits than were the average heritability estimates of the items building them up. Because the composite traits also can be expected to be more stable over time and between situations than the individual items, the former should be considered for use as selection traits. When defining composite traits for breeding purposes it is advised, in order to improve heritability, to investigate if genetic parameters should be considered rather than aggregating items based on a principal component analysis or a factor analysis performed on phenotypic data alone. The results also indicate that from a heritability perspective, behavioural measurements should be objective rather than subjective, and neutral rather than passing value judgments.

Dog Mentality Assessment is possible and recommended to use for selection of breeding animals with the goal to improve several everyday life

behavioural traits, for example non-social fear, in the Swedish Rough Collie population. The summated scales method to compute Dog Mentality Assessment personality trait scores seems to perform at least as well as the factor scores method; estimated heritabilities and genetic correlations between Dog Mentality Assessment results and everyday life behaviour as described by dog owners are generally equal or greater for the summated scales. Because they are also easier to use for practical breeding purposes, summated scales are the first choice in a breeding program for Rough Collies based on Dog Mentality Assessment data.

Dog populations often are small, and would benefit in terms of accuracy of selection, selection intensity and decreased inbreeding rate, from utilizing data from related populations in a joint genetic evaluation. Therefore, collaboration between populations should be encouraged. Because an effective utilization of data from a related population requires some form of genetic evaluation, this emphasizes the need of a more extensive use of BLUP in dog breeding.

# 7 Future challenges

One benefit of introducing effective methods to select breeding animals is obvious and has already been discussed; higher accuracy of selection enables faster genetic progress. Furthermore, a joint genetic evaluation where several populations are included, for example the same breed in two different countries, may increase selection intensity. Other examples of benefits are the possibility to estimate genetic correlations between traits, and to calculate genetic trends to study if the breeding is successful.

Naturally, introduction of new methods also comes with challenges. Estimating BLUP breeding values requires a high level of expertise, which typically is connected with an economic investment. When breeding is practiced under commercial conditions, as is the case for production animals such as cattle or poultry, an investment in expertise is easy to defend as long as it pays off in terms of increased productivity. For a typical dog breeder, the value of genetically improved dogs is not as easily translated into money, and the benefit of an economic investment in expertise may therefore be less obvious. Because most dog breeders produce only few litters each year (in many cases even less than one on average), a single breeder alone cannot be expected to make this sort of investment. Instead this has to be a joint effort by many breeders, for example within already existing breed or kennel organisations. Exceptions exist, for example the Swedish Armed Forces' breeding program, which produces dogs in a comparably large scale – approximately 300 puppies each year – and recruits replacement breeding animals almost entirely among the dogs born within the program. Under these conditions, it is recommended to start using EBVs for selection, even if it means that the Swedish Armed Forces needs to build their own infrastructure.

No matter how good tools the breeders are provided with, no genetic progress will happen unless the tools are widely utilized. Among dog breeders in general, selection on EBVs rather than on phenotypes is still a fairly novel

concept. Unless breeders trust EBVs as a basis for selection and understand why they are to be preferred, they cannot be expected to actually use them. For example, it might not be immediately obvious that a phenotypically good dog can have a lower breeding value than a phenotypically less good dog, or that existing breeding restrictions based on phenotypes should be avoided and breeding decisions instead be based on EBVs. It may also be confusing that a dog's breeding value is not necessarily static (when new information is added) and even can be expected to decrease with time (if there is a genetic progress and the base for calculating EBVs is updated, that is). Therefore, to be successful in implementing BLUP, education and information are crucial. As an example, the studies presented in Paper IV constitute the basis for a breeding program for temperament in Rough Collie in Sweden. Since March 2012, BLUP breeding values for, e.g., DMA Curiosity/Fearlessness and Gunshot avoidance are published openly on the Swedish Collie Club web page (www.svenskacollieklubben.se). The Swedish Collie Club, supported by the Swedish University of Agricultural Sciences, works with informing and educating breeders and puppy buyers about breeding and temperament, how to interpret and use breeding values, etc. In the near future, a system for certifying litters with good chances of becoming less fearful than Rough Collies in general will be launched. A breeder with a litter for sale, where the average breeding value of the parents for Curiosity/Fearlessness (and maybe also for Gunshot avoidance) is better than the breed average, will be allowed to market the litter as "Mentally Sound Collie Certified by the Swedish Collie Club". The goal is that with time it will become natural for puppy buyers to request only certified puppies, thereby encouraging the breeders to breed for less fearful Rough Collies.

To achieve genetic progress, it is necessary for the breeders to agree on a breeding goal. Defining a breeding goal for behavioural traits can be expected to be challenging. One reason among many is that behavioural traits often do not have an obvious optimum. If, for example, breeding for hip dysplasia, the less dysplastic hips the better. If breeding for Sociability, Chase-proneness or Playfulness it is more complicated. How to find (and agree on) an optimum for a behavioural trait is therefore a very important area for future research. Furthermore, breeding goals are likely to differ between breeds. Owing to the large number of dog breeds and limited resources for most breed clubs, it would be desirable not only to define a breeding goal for one or a couple of breeds, but also to develop relatively simple methods for defining the breeding goal, which can then be managed by the dog organisations.

Dog breeding is not all about the behaviours measured in the measurement methods analysed in this thesis. First, there are most likely other behavioural

traits of importance in addition to the ones studied (and vice versa; not all of the studied traits are necessarily important to breed for). Second, there are other types of traits, primarily health-related, that should be considered. It is important to make clear what traits are important for a certain breed, and then select for these rather than breeding for something only because it is being measured (and is possible to estimate breeding values for). In this context it is relevant to emphasize the importance of putting much more weight on behaviour (and health) than on appearance when selecting breeding animals. In a simulation study on the Finnish Rottweiler population, Mäki *et al.* (2005) compared genetic responses for hip and elbow dysplasia, behaviour and appearance when different breeding schemes were applied. They concluded that to achieve genetic improvement for health and behavioural traits, changes were required in the current dog breeding programs; emphasis on selection for health and behaviour had to be increased at the expense of appearance.

In all studies in the thesis, the quality of a measurement from a selection perspective has been established partly by interpreting genetic parameter estimates in terms of expected genetic progress. Conclusions based on these interpretations thus assume that the estimates are correct. As an alternative, it would be interesting to use cross-validation. For example, different methods to compute composite trait scores could be compared by using a part of the population for computing scores and estimating EBVs for each method, and then predicting outcome on the remaining part of the population. The predicted outcome can then be compared with the actual phenotypes.

# 8   Avelsvärdering för beteende hos hund

## 8.1   Bakgrund

Hundars beteendeegenskaper är viktiga av flera skäl. Till exempel kan synskadade, poliser, jägare och lantbrukare ha stor nytta av ledar-, narkotika-, jakt- och vallhundar för att lösa angelägna uppgifter. För att detta ska fungera måste hundarna vara mentalt rustade för dessa ändamål. Samma sak gäller vanliga familjehundar. De ska helst kunna fungera både i staden och på landet, kunna lämnas ensamma hemma åtminstone några timmar utan att bita sönder inredningen, åka bil, buss och tåg, inte bli rädda i onödan eller arga i fel situationer, kunna fungera tillsammans med andra djur och främmande människor, och så vidare. En hunds beteende påverkas av många olika faktorer. Det är väl belagt att mentalitet eller personlighet, och även funktionsegenskaper av mer specifik betydelse för exempelvis jakt- eller vallhundar, delvis styrs av gener.

Hundars beteendeegenskaper är alltså både viktiga och ärftliga. Därmed är de även angelägna att ta hänsyn till i avelsarbetet. Hundavel bedrivs i stor utsträckning inom raser. I de stora hundorganisationerna definieras rasbegreppet utifrån släktskap. Som rasren räknas den hund som är registrerad i en stambokförande hundorganisation, och för att bli registrerad krävs i allmänhet att båda föräldrarna är registrerade. Därigenom ska säkerställas att alla hundar i rasen härstammar från de hundar som anses som rasens grundare. Tillsammans med det sätt på vilket hundavel traditionellt bedrivits, med systematisk inavel, omfattande användning av enskilda hanhundar och stark selektion, har detta lett till att den genetiska variationen inom raser är liten. Hunden är vårt äldsta husdjur och domesticerades för minst 15 000 år sedan. Det beskrivna rasbegreppet är betydligt yngre och började i princip inte tillämpas i större skala förrän under 1800-talet. Sedan dess har det även skett en förskjutning i avelsmål från att tidigare främst ha handlat om praktisk funktion – för vilken beteendeegenskaper är av avgörande betydelse – till att

fokusera på utseendemässig överensstämmelse med den så kallade rasstandarden som från början formulerades i samband med att rasen skapades.

För så kallat kvalitativa egenskaper, det vill säga egenskaper som styrs av enskilda gener, kan man ibland med blotta ögat eller andra enkla metoder bestämma hundens genotyp. Det kan exempelvis handla om vissa sjukdomar eller pälsfärg. För en del kvalitativa egenskaper finns också möjligheten att göra ett gentest. Beteendeegenskaper däremot är kvantitativa vilket innebär att de styrs av ett stort antal gener och dessutom påverkas av miljöfaktorer, som uppväxtmiljö och den omgivning och hantering individen utsätts för i vardagen. Kombinationen av gener och inverkan av omgivning påverkar hur individen reagerar och beter sig i olika situationer. Detta betyder att när man ska selektera avelsdjur kan det vara svårt att bedöma om en hund är bra för en viss egenskap på grund av att den har bra genotyp för egenskapen, eller om den är bra för att den haft gynnsamma förutsättningar i övrigt, till exempel en bra ägare.

Inom hundaveln praktiseras nästan uteslutande så kallad fenotypselektion, vilket betyder att avelsdjur rekryteras huvudsakligen på grundval av sina egna prestationer. Eftersom fenotypen för en kvantitativ egenskap inte är ett säkert mått på vad hunden nedärver om den används i avel är alltså risken stor att man vid fenotypselektion inte lyckas finna de djur som verkligen är de bästa avelsdjuren. Inom modern husdjursavel för livsmedelsproducerande djur som nötkreatur, grisar eller värphöns, och även inom ridhästaveln, används BLUP-metoden för att skatta så kallade avelsvärden för de egenskaper man vill avla för (BLUP står för Best Linear Unbiased Prediction, vilket är en beskrivning av metodens egenskaper). Metoden är väl och framgångsrikt beprövad. Det finns några exempel där BLUP används inom hundavel, till exempel för höft- och armbågsledsdysplasi i flera länder och för jaktegenskaper hos vorsteh i Norge, hos finsk spets i Finland och hos drever i Sverige. Sett i ett större perspektiv är detta dock fortfarande undantag.

Poängen med avelsvärden är att de är säkrare mått på vad en hund kommer att nedärva än vad hundens eget resultat är. Därigenom öppnas möjligheten att göra snabbare framsteg i avelsarbetet. Den ökade säkerheten kommer sig av att man då BLUP-avelsvärden beräknas dels väger in alla släktingars resultat, dels korrigerar hundens eget resultat för en egenskap för olika miljöfaktorer. Om det exempelvis är så att äldre hundar i allmänhet presterar bättre för en egenskap så "höjer man upp"/korrigerar de unga hundarnas resultat så att alla hundar blir jämförbara oavsett ålder. Samma sak om tikar och hanar är olika, och så vidare. Skälet till att man väger in släktingars resultat är helt enkelt att släktingar har en viss andel gener gemensamma. Om exempelvis en hund i en kull är mycket bra för någon egenskap medan resterande sju är dåliga, kan man

misstänka att den bra hundens gener kanske inte är så bra som dess egna prestationer först ger anledning att tro. I så fall är den hunden kanske inte heller lysande som avelsdjur.

När man beräknar en hunds BLUP-avelsvärde utgår man från hundens (och dess släktingars) fenotyp. Själva måttet är alltså fortfarande lika viktigt som vid fenotypselektion. Beteendeegenskaper kan mätas på många olika sätt, och det är inte självklart vilket som är att föredra. Exempelvis är konkurrensbaserade mått inte alla gånger idealiska att behandla statistiskt på det sätt som är nödvändigt. Det finns flera orsaker till varför, en av dem är att om man bara bedömer en hund på en skala från bra till dålig (som man ju ofta gör i provsammanhang) så vet man inte i efterhand på vilket sätt den dåliga hunden var dålig. Var den för lugn eller för intensiv, för försiktig eller övermodig och så vidare. Och i en situation där man vill väga samman flera släktingars resultat är det viktigt att veta *hur* respektive individ brister, inte bara *att* den brister.

Det övergripande syftet med denna avhandling var att studera förutsättningarna för att förbättra hundars beteendeegenskaper genom avel. Därför analyserades ett antal metoder som används eller har använts för att mäta olika beteendeegenskaper hos hundar, med syftet att dels undersöka hur väl respektive metod skulle lämpa sig för att användas i avel för att åstadkomma ett genetiskt framsteg, dels studera vad som kännetecknar en bra mätmetod. Mer specifikt var syftet att:

➢ Skatta genetiska parametrar baserat på åtta olika metoder att mäta mentala egenskaper och jakt- och vallningsegenskaper. Exempel på genetiska parametrar är arvbarhet och genetisk korrelation. Arvbarheten för en egenskap är ett statistiskt mått på hur mycket av den uppmätta variationen för egenskapen som beror på genetiska skillnader mellan individer. Ju högre arvbarhet, desto enklare är egenskapen att förändra genom avel. Den genetiska korrelationen beskriver hur starkt två egenskaper hänger ihop genetiskt; om den genetiska korrelationen är hög så betyder det att en genetisk förändring i den ena egenskapen (exempelvis som en följd av systematisk avel) kommer medföra en genetisk förändring även i den andra egenskapen.
➢ Studera hur graden av objektivitet och neutralitet vid beteendemätningar påverkar arvbarheten.
➢ Undersöka hur olika metoder att definiera och beräkna underliggande egenskaper påverkar arvbarheten för dessa och/eller deras genetiska korrelation till egenskaper i avelsmålet. Definitionerna av underliggande egenskaper utgick från så kallad principalkomponentanalys eller faktorsanalys. Dessa analysmetoder används för att med utgångspunkt från

hur ett antal variabler är korrelerade till varandra identifiera ett lägre antal underliggande egenskaper.

➢ Studera hur en gemensam avelsvärdering för två länder påverkar avelsvärdenas säkerhet, jämfört med om avelsvärderingen görs inom land.

➢ Skatta genetiska korrelationer mellan beteende i ett temperamentstest och i vardagen så som det uppfattas av hundägarna själva.

## 8.2  Sammanfattning av studierna

De beteendedata som analyserades kom från två olika versioner av Svenska Vallhundsklubbens Arbetsbeskrivning (border collie), svenska och norska jaktprov för engelsk setter, två typer av protokoll som används för att registrera beteendet hos schäfrar som genomför Försvarsmaktens lämplighetstest, Svenska Brukshundklubbens Mentalbeskrivning Hund (långhårscollie), och en utökad version av enkäten C-BARQ (Canine Behavioral Assessment and Research Questionnaire) (långhårscollie). Enkätdata samlades in som en del i studien, alla övriga beteendemått hade samlats in av ansvarig ras- eller specialklubb, respektive av Försvarsmakten. Vid Arbetsbeskrivningen mäts olika vallningsegenskaper, vid jaktproven jaktegenskaper av vikt för stående fågelhundar, vid Mentalbeskrivning Hund och Försvarsmaktens lämplighetstest egenskaper som bedömts som viktiga för bruks- och tjänstehundar, och i enkäten hundens beteende i vardagen.

De genomsnittliga arvbarheterna för alla måtten inom respektive mätmetod varierade från 0,10 (svenska jaktprov för engelsk setter) till 0,32 (version 1 av Arbetsbeskrivningen), vilket i stort överensstämmer väl med tidigare studier av andra mätmetoder. För fyra av mätmetoderna användes beteendemåtten för att beräkna sammanfogade mått för underliggande egenskaper. Arvbarhets-skattningarna för de underliggande egenskaperna varierade mellan 0,16 och 0,20 och var högre än för de ursprungsmått som använts för att beräkna dem. För alla mätmetoder påverkades en majoritet av måtten av olika miljöeffekter.

Sammanfogade mått från Mentalbeskrivning Hund visade starka genetiska korrelationer till olika vardagsegenskaper så som de beskrivits av hundägare i enkäten, till exempel mentalbeskrivningsegenskapen Nyfikenhet/Orädsla med enkätegenskapen Icke-social rädsla (-0,70) och mentalbeskrivningsegenskapen Jaktlust med enkätegenskapen Jakt (0,73). Detta betyder att svenska collieuppfödare kan välja avelsdjur med utgångspunkt från Mentalbeskrivning Hund för att åstadkomma en genetisk förändring i viktiga vardagsegenskaper, exempelvis Icke-social rädsla.

Vad gäller hur objektivitet och neutralitet påverkar arvbarheten så tyder vissa resultat på att en mer objektiv och neutral mätning är att föredra, och inga

resultat indikerar motsatsen. Sammantaget tycks därför objektiva och neutrala skalor vara att rekommendera.

När underliggande egenskaper definieras kan man utgå från hur mätvärdena för olika mått korrelerar till varandra rent fenotypiskt. I ett avelsarbete är det emellertid de genetiska korrelationerna som är intressantast, och genetiska korrelationer kan skilja sig från de fenotypiska. När Försvarsmaktens lämplighetstest studerades tydde resultaten på att man borde ta hänsyn även till genetiska parametrar när de underliggande egenskaperna definierades eftersom detta ledde till generellt högre arvbarhetsskattningar. Analyserna av Mentalbeskrivning Hund visade att även en relativt enkel metod att beräkna mått för underliggande egenskaper gav minst lika höga arvbarheter som en mer komplicerad metod, och att de genetiska korrelationerna till avelsmåls-egenskaperna som mättes med enkäten var desamma mellan den enklare och den mer avancerade metoden.

Analyserna av jaktprovsdata för engelsk setter visade att svenska uppfödare kan nå 66% snabbare genetiskt framsteg (i genomsnitt för de analyserade jaktegenskaperna) om de använder sig av BLUP-avelsvärden som skattats med utgångspunkt enkom från svenska data när de väljer avelsdjur i stället för om de utgår endast från de potentiella avelsdjurens egna resultat. Om de dessutom inkluderar norska prov- och härstamningsdata i avelsvärderingen kan framsteget ökas ytterligare. Totalt sett är skillnaden i möjligt genetiskt framsteg nästan en fördubbling (95%) om man väljer avelsdjur utifrån BLUP-avelsvärden från en gemensam norsk-svensk avelsvärdering, jämfört med om man utgår endast från hundarnas egna provresultat. Norska uppfödare kan nå 87% snabbare genetiskt framsteg (i genomsnitt för de sex egenskaperna) om de använder sig av "norska" BLUP-avelsvärden när de väljer avelsdjur i stället för att utgå från hundarnas provresultat. Däremot blir den ytterligare ökningen om de inkluderar även svenska data marginell, endast 1%. Dock finns en betydande vinst med en gemensam avelsvärdering på ett annat sätt. En gemensam avelsvärdering innebär nämligen att det blir enklare att jämföra hundar mellan länderna. En norsk uppfödare får alltså relevant information om de svenska hundarna vilket betyder att det blir fler tänkbara avelsdjur att välja bland. Samma sak gäller givetvis omvänt för de svenska uppfödarna. Därigenom kan de inom hundavel mycket vanliga problemen relaterade till små populationer minskas. Ytterligare ett sätt att hantera dessa problem vore att tillämpa mindre strikta barriärer mellan snarlika raser eller rasvarianter.

## 8.3  Kortfattade slutsatser

Samtliga studerade mätmetoder visar tillräckligt hög genetisk variation för merparten av måtten för att metoderna ska vara möjliga att använda i ett avelsarbete. Eftersom arvbarheterna som regel är låga eller medelhöga, och på grund av att måtten påverkas av olika systematiska miljöfaktorer, bör man sträva efter att införa rutinmässig skattning av BLUP-avelsvärden att använda för selektion för avel. Därigenom blir det möjligt att med större säkerhet välja de avelsdjur som kommer generera ett genetiskt framsteg. Skattning av avelsvärden har redan införts för collie. Eftersom mätegenskaperna (från Mentalbeskrivning Hund) är starkt genetiskt kopplade till viktiga avelsmålsegenskaper (beteende i vardagen) bör man fortsätta på den inslagna vägen och verka för att avelsvärdena verkligen nyttjas av uppfödarna som grund för selektion av avelsdjur. Hundraser är ofta numerärt små, och därför är internationellt samarbete mycket viktigt. Genom en gemensam avelsvärdering inkluderande hundar från två eller fler länder kan både avelsvärdenas säkerhet och antalet selektionskandidater ökas, vilket i sin tur underlättar avelsarbetet betydligt.

# References

Arnott, E.R., Early, J.B., Wade, C.M. and McGreevy, P.D. (2014). Estimating the economic value of Australian stock herding dogs. *Animal Welfare* 23, 189-197.

Axelsson, E., Ratnakumar, A., Arendt, M.-J., Maqbool, K., Webster, M.T., Perloski, M., Liberg, O., Arnemo, J.M., Hedhammar, Å. and Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. *Nature* 495, 360-365.

Brenøe, U. T., Larsgard, A. G., Johannessen, K.-R. and Uldal, S. H. (2002). Estimates of genetic parameters for hunting performance traits in three breeds of gun hunting dogs in Norway. *Applied Animal Behaviour Science* 77, 209-215.

Calus, M.P.L. (2010). Genomic breeding value prediction: methods and procedures. *Animal* 4, 157-164.

Cederström, A., Johansson, A. and Andersson, K. (1994). *Boken om drever*. Setterns förlag, pp. 121-135.

Clutton-Brock, J. (1995). Origins of the dog: domestication and early history. In: Serpell, J. (ed) *The domestic dog – its evolution, behaviour and interactions with people*. Cambridge University Press, Cambridge, pp. 51-64.

Correau, J. F. and Langlois, B. (2005). Genetic parameters and environmental effects which characterise the defence ability of the Belgian shepherd dog. *Applied Animal Behaviour Science* 91, 233-245.

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

Egenvall, A., Hedhammar, Å., Bonnett, B.N. and Olson, P. (1999). Survey of the Swedish Dog Population: Age, Gender, Breed, Location and Enrolment in Animal insurance. *Acta Veterinaria Scandinavica* 40, 231-240.

Falconer, D.S. and Mackay, T.F.C. (1996). *Introduction to Quantitative Genetics*. 4th ed. Pearson Education Limited, Harlow, UK.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136, 245-257.

Goddard, M.E. and Beilharz, R.G. (1982). Genetic and Environmental Factors Affecting the Suitability of Dogs as Guide Dogs for the Blind. *Theoretical and Applied Genetics* 62, 97-102.

Hair, J.F., Anderson, R.E., Tatham R.L. and Black, W.C. (1998). *Multivariate Data Analysis*. 5th ed. Prentice-Hall, Upper Saddle River, New Jersey. Chapter 3.

Hayes, B.J, Lewin, H.A. and Goddard, M.E. (2013). The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends in Genetics* 29, 206-214.

Henderson,C.R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32, 69-83.

Hoffmann, U., Hamann, H. and Distl, O. (2002). Genetische Analyse von Merkmalen der Leistungsprüfung für Koppelgebrauchshunde. 1. Mitteilung: Leistungsmerkmale. Berl. *Münch. Tierärzl. Wschr*. 116, 81-89.

Hsu, Y.Y. and Serpell, J.A. (2003). Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association* 223, 1293-1300.

Hubrecht, R. (1995). The welfare of dogs in human care. In: Serpell, J. (ed) *The domestic dog – its evolution, behaviour and interactions with people*. Cambridge University Press, Cambridge, pp. 51-64.

Hundansvarsutredningen (2003). *Hund i rätta händer – om hundägarens ansvar*. Statens offentliga utredningar SOU 2003:46. Fritzes, Stockholm. (Swedish governmental inquiry.)

Häggström, J., Hansson, K., Kvart, C. and Swenson, L. (1992). Chronic valvular disease in the Cavalier King Charles Spaniel in Sweden. *Veterinary Record* 131, 549-553.

Karjalainen, L., Ojala, M. and Vilva, V. (1996). Environmental effects and genetic parameters for measurements of hunting performance in the Finnish Spitz. *Journal of Animal Breeding and Genetics* 113, 525-534.

Larson, G. and Bradley, D.G. (2014). How Much Is That in Dog Years? The Advent of Canine Population Genomics. *PLOS Genetics* 10, 1-3.

Larson, G., Karlsson, E.K., Perri, A., Webster, M.T., Ho, S.Y.W., Peters, J., Stahl, P.W., Piper, P.J., Lingaas, F., Fredholm, M., *et al.* (2012). Rethinking dog domestication by integrating genetics, archeology, and biogeography. *PNAS* 109, 8878-8883.

Liimatainen, R., Liinamo, A.-E. and Ojala, M. (2008). Genetic factors affecting behavior test results in Rottweilers. *Journal of Veterinary Behavior: Clinical Applications and Research* 3, 178.

Liinamo, A.-E. (2004). Genetic trends in hunting behaviour in the Finnish Hound. In: *Book of Abstracts of the 55th Annual Meeting of the European Association for Animal Production*. Slovenia, 5-9 September 2004. Wageningen Academic Publishers.

Liinamo, A.-E. and van Arendonk, J.A.M. (2006). Genetic parameters of show quality and its relationships with working traits and hip dysplasia in Finnish Hounds. *8th World Congress on Genetics Applied to Livestock Production*, August 13-18, 2006, Belo Horizonte, MG, Brasil, Paper 10-10.

Liinamo, A.-E., van den Berg, L., Leegwater, P. A. J., Schilder, M. B. H., van Arendonk, J. A. M. and van Oost. B. A. (2007). Genetic variation in aggression-related traits in Golden Retriever dogs. *Applied Animal Behaviour Science* 104, 95-106.

Lindberg, S., Strandberg, E. and Swenson, L. (2004). Genetic analysis of hunting behavior in Swedish Flatcoated Retrievers. *Applied Animal Behaviour Science* 88, 289-298.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., Zody, M.C., *et al.* (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819.

MacIsaac, P., Blondin, J. and Sawyer, B. (2005). The Royal Canadian Mounted Police – Police Dog Training Centre. In: *Proceedings from the 4th International Working Dog Conference*, 23-27 January 2005, pp. 13-19.

Mackenzie, S.A., Oltenacu, E.A.B. and Leighton, E. (1985). Heritability estimate for temperament scores in German Shepherd Dogs and its genetic correlation with hip dysplasia. *Behavior Genetics* 15, 475-482.

Madsen, P. and Jensen. J. (2010). *A User´s Guide to DMU – A Package for Analysing Multivariate Mixed Models*. Version 6, release 5.0. University of Aarhus, Faculty Agricultural Sciences (DJF), Dept. of Genetics and Biotechnology, Research Centre Foulum, Tjele, Denmark.

Malm, S., Fikse, W.F., Danell, B. and Strandberg, E. (2008). Genetic variation and genetic trends in hip and elbow dysplasia in Swedish Rottweiler and Bernese Mountain Dog. *Journal of Animal Breeding and Genetics* 125, 403–412.

Malm, S., Sørensen, A.C., Fikse,W.F. and Strandberg, E. (2013). Efficient selection against categorically scored hip dysplasia in dogs is possible using best linear unbiased prediction and optimum contribution selection: a simulation study. *Journal of Animal Breeding and Genetics* 130, 154-164.

Marshall-Pescini, S. and Kaminski, J. (2014). The Social Dog: History and Evolution. In: Kaminski, J. and Marshall-Pescini, S. (eds) *The Social Dog: Behaviour and Cognition*. Academic Press (Elsevier), San Diego, pp. 3-32.

McGreevy, P. (2008). Comment: We must breed happier, healthier dogs. New Scientist. 2677, 18.

McGreevy, P.D. and Nicholas, F.W. (1999). Some practical solutions to welfare problems in dog breeding. *Animal Welfare* 8, 329-341.

Meuwissen, M.E. (1997). Maximizing the response of selection with a predefined rate of inbreeding. *Journal of Animal Science* 75, 934-940.

Meyer, F., Schwalder, P., Gaillard, C. and Dolf, G. (2012). Estimation of genetic parameters for behavior based on results of German Shepherd Dogs in Switzerland. *Applied Animal Behaviour Science* 140, 53-61.

Murphy, J.A. (1997). Describing categories of temperament in potential guide dogs for the blind. *Applied Animal Behaviour Science* 58, 163-178.

Mäki, K. (2004). *Breeding against hip and elbow dysplasia in dogs*. Diss. University of Helsinki. Finland.

Mäki, K., Groen, A.F., Liinamo, A.-E. and Ojala, M. (2002). Genetic variances, trends and mode of inheritance for hip and elbow dysplasia in Finnish dog populations. *Animal Science* 75, 197–207.

Mäki, K., Liinamo, A.-E., Groen, A.F., Bijma, P. and Ojala, M. (2005). The effect of breeding schemes on the genetic response of canine hip dysplasia, elbow dysplasia, behaviour traits and appearance. *Animal Welfare* 14, 117-124.

Olsen, H.F., Klemetsdal, G., Odegard, J. and Arnason, T. (2012). Validation of alternative models in genetic evaluation of racing performance in North Swedish and Norwegian cold-blooded trotters. *Journal of Animal Breeding and Genetics* 129, 164-170.

Ostrander, E.A. and Wayne, R.K. (2005). The canine genome. *Genome Research* 15, 1706-1716.

Parker, H.G., Kim, L.V., Sutter, N.B., Carlson, S., Lorentzen, T.D., Malek, T.B., Johnson, G.S., DeFrance, H.B., Ostrander, E.A. and Kruglyak, L. (2004). Genetic Structure of the Purebred Domestic Dog. *Science* 304, 1160-1164.

Pösö, J., Pedersen, J., Lidauer, M., Mäntysaari, E.A., Strandén, I., Madsen, P., Nielsen, U.S., Eriksson, J.-Å., Johansson, K., Aamand, G.P. (2006). Joint Nordic Test Day Model: Experiences with the New Model. *Interbull Open Meeting*, Kuopio, Finland, June 4th – 6th.

Rogers, G.W., Banos, G. and Sander-Nielsen, U. (1999). Genetic correlations among protein yield, productive life, and type traits from the United States and diseases other than mastitis from Denmark and Sweden. *Journal of Dairy Science* 82, 1331-1338.

Rooney, N. J. (2009). The welfare of pedigree dogs: Cause for concern. *Journal of Veterinary Behavior: Clinical Applications and Research* 4, 180-186. (Editorial)

Ruefenacht, S., Gebhardt-Henrich, S., Miyake, T. and Gaillard, C., (2002). A behaviour test on German Shepherd dogs: heritability of seven different traits. *Applied Animal Behaviour Science* 79, 113-132.

Saetre, P., Strandberg, E., Sundgren, P.-E., Pettersson, U., Jazin, E. and Bergström, T.F. (2006). The genetic contribution to canine personality. *Genes, Brain and Behavior* 5, 240-248.

SAS (2008). Release 9.2. SAS Institute Inc., Cary, NC, USA.

Schiefelbein, K.M. (2012). *Estimation of genetic parameters for behavioral assessment scores in Labrador Retrievers, German Shepherd Dogs, and Golden Retrievers*. MSc thesis. Kansas State University.

Schiefelbein, K.M. (2013). Estimation of genetic parameters for behavioral assessment scores in Labrador Retrievers, German Shepherd Dogs, and Golden Retrievers. In: *Abstracts from the 8th International Working Dog Conference*, San Antonio, Texas. p. 6.

Serpell, J.A. and Duffy, D.L. (2014). Dog Breeds and Their Behavior. In: Horowitz, A. (ed) *Domestic Dog Cognition and Behavior*. Springer-Verlag, Berlin Heidelberg, pp 31-57.

Slabbert, H. (2008). The selection, breeding and training of working dogs. In: *Speaker's abstracts from the Odour Detection by Animals conference*, 16-20 June 2008, Norway.

Stock, K.F. and Distl, O. (2010). Simulation study on the effects of excluding offspring information for genetic evaluation versus using genomic markers for selection in dog breeding. *Journal of Animal Breeding and Genetics* 127, 42-52.

Strandberg, E., Jacobsson, J. and Saetre, P. (2005). Direct genetic, maternal and litter effects on behaviour in German shepherd dogs in Sweden. Livestock Production Science 93, 33-42.

Svartberg, K. (2002). Shyness-boldness predicts performance in working dogs. *Applied Animal Behaviour Science* 79, 157-174.

Svartberg, K. (2005). A comparison of behaviour in test and in everyday life: evidence of three consistent boldness-related personality traits in dogs. *Applied Animal Behaviour Science* 91, 103-128.

Svartberg, K. (2006). Breed-typical behaviour in dogs – Historical remnants or recent constructs? *Applied Animal Behaviour Science* 96, 293-313.

Svartberg, K. (2013). *Utvärdering av Beteende- och Personlighetsbeskrivning Hund – första året med BPH*. Report for the Swedish Kennel Club. [online] http://www.skk.se/Global/Dokument/Om-SKK/BPH/Forsta_aret_med_BPH_inkl_appendix.pdf [2014-10-09]

Swedish Kennel Club (2014). *Index för HD och ED*. [online] http://www.skk.se/uppfodning/halsa/halsoprogram/index-for-hd-och-ed/ [2014-10-02]

Swenson, L. (1983). *Vallprover som urvalsinstrument vid avel för bättre vallhundar*. Report from the Swedish Kennel Club project "Hundavel".

Tjänstehundsutredningen (2005). *Hundgöra – att göra hundar som gör nytta*. Statens offentliga utredningar SOU 2005:75. Fritzes, Stockholm. (Swedish governmental inquiry.)

Van der Waaij, E.H., Wilsson, E. and Strandberg, E. (2008). Genetic analysis of results of a Swedish behavior test on German Shepherd Dogs and Labrador Retrievers. *Journal of Animal Science* 86, 2853-2861.

Van Rooy, D., Arnott, E.R., Early, J.B., McGreevy, P. and Wade, C. (2014). Holding back the genes: limitations of research into canine behavioural genetics. *Canine Genetics and Epidemiology* 1, 1-11.

Vanderloo, J. (2005). Breeding and Developing Detector Dogs. In: *Proceedings from the 4th International Working Dog Conference*, 23-27 January 2005, pp. 151-153.

Vangen, O. and Klemetsdal, G. (1988). Genetic studies of Finnish and Norwegian test results in two breeds of hunting dogs. In: *Proceedings of the Sixth World Conference on Animal Production*. Helsingfors, Finland, p. 496.

Vazire, S., Gosling, S.D., Dickey, A.S. and Schaprio, S.J. (2007). Measuring personality in nonhuman animals. In: Robins, R.W., Fraley, R.C., Krueger, R. (eds) *Handbook of Research Methods in Personality Psychology*. Guilford, New York, pp. 190–206.

Viklund, Å. (2010). *Genetic evaluation of Swedish Warmblood horses*. Diss. Uppsala: Swedish University of Agricultural Sciences.

Willis, M. B. (1995). Genetic aspects of dog behaviour with particular reference to working ability. In: Serpell, J. (ed) *The domestic dog – its evolution, behaviour and interactions with people*. Cambridge University Press, Cambridge, pp. 51-64.

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

Wilsson, E. and Sundgren, P.-E. (1997). The use of a behaviour test for selection of dogs for service and breeding. II. Heritability for tested parameters and effect of selection based on service dog characteristics. *Applied Animal Behaviour Science* 54, 235-241.

# Acknowledgements

The work of this thesis was performed at the Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences (SLU). The project was primarily funded by the department. The Swedish Collie Club and the Swedish Kennel Club also contributed with financial support. Travel grants from the SLU fund for internationalization of postgraduate studies enabled study visits and participation in conferences in the U.S. and in Canada. All financial support is greatly acknowledged. The Norwegian English Setter Club, the Norwegian Kennel Club, the Swedish Armed Forces, the Swedish Kennel Club, the Swedish Setter Club for English Setter and the Swedish Sheepdog Society are acknowledged for permission to use their data.

Many people have contributed in different ways to this thesis, and I am sincerely grateful to every one of you! I want to express my special thanks to:

My main supervisor Erling Strandberg and my co-supervisors Katja Nilsson and Kenth Svartberg for your extensive knowledge, for answering my questions and for allowing me to do things my way, and my co-supervisor Freddy Fikse for sharing your extensive knowledge, for always(!) having or finding out a solution to any scientific problem related to animal breeding, and for *not* always allowing me to do things my way.

Gunnar Klemetsdal who did not only welcome me to spend half a year at the Norwegian University of Life Sciences, but also was an excellent unofficial supervisor and a supportive co-author.

All nice and extremely engaged dog breeders, dog owners, people from the dog organizations and within the working dog community, who shared their time, knowledge, experience and commitment by answering my questions and discussing the realities of dog breeding: Gunvor af Klinteberg-Järverud, Curt Blixt, Leif Carlsson, Lasse Eriksson, Helena Frögéli, Erik Wilsson…and many more. I did not always agree with you, but the discussions were always great fun and worthwhile, and I hope they will continue!

Previous and present colleagues at the department, especially the amazingly clever and hard-working PhD students! ☺

Barbara Havlena, Stewart Hilliard, Jane Russenberger, Brenda Sawyer and Curtis Shull who made my study visits most valuable experiences.

Bertil Norbelie for helping me to see the working dog world from a different perspective, and for your generous efforts with the working dog council. And for all the shrimp sandwich lunches, I appreciated them a lot!

Sven, Per, Ann-Sofie, Susanne, Benny, Göran and all other dog trainer colleagues of mine from the 90-ies; I still have great use of things you said and did and that I learned from working together with you!

Finally, the most important ones, my friends for being my friends, my family for being my family, and Suzana for being exactly who you are!