Social Behavior and Cognitive Bias in the Domestic Dog (*Canis familiaris*)

Carolyn J. Loyer

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Michael D. Beecher, Chair

Robin L. Foster

James C. Ha

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Social Behavior and Cognitive Bias in the Domestic Dog (*Canis familiaris*)

Carolyn J. Loyer

Chair of the Supervisory Committee:
Michael D. Beecher
Psychology

Investigation of individual differences in behavior has increased in recent decades, with implications for both applied and basic science. Research with species that have evolved alongside humans presents an opportunity to improve understanding and welfare simultaneously. This dissertation first examines the potential applications of personality research with nonhuman animals of any species, before shifting focus to a species that is ubiquitous in the lives of humans around the world; domestic dogs (*Canis familiaris*).

I discuss the development and validation of a pictorial questionnaire designed to measure three dimensions of canine body language during interactions with conspecifics: Aggression, Arousal, and Boldness. Results indicate that these three dimensions are correlated, with body language identified as higher arousal associated with being more aggressive and more shy/fearful when rated by experts as well as when reported by owners. Observer ratings of dogs' body

language during an introduction to a live dog correlated with owner responses on the Arousal and Aggression dimensions, but not Boldness.

Subsequently, I compare dog behavior in a live dog introduction to three other conditions (control, a novel object, and a model dog) to evaluate the validity of using a model dog as a proxy for a live dog when assessing dog-dog sociability. In general, dogs treated the model dog as intermediate between the live dog and a novel object. Dogs' scores on Aggression in an owner survey correlated with their behavior in the experiment. Dogs that were higher in aggression stared longer and oriented their bodies more directly toward the model dog and live dog compared to the novel object and control conditions. Dogs with lower Aggression scores did not treat the conditions differently in this regard.

Finally, I investigate potential cognitive underpinnings of canine aggression toward conspecifics. I compared dogs' judgement bias (the tendency to treat an ambiguous stimulus as predicting a positive vs. negative outcome) in a spatial discrimination task to individual differences assessed by owner survey, including aggression toward other dogs and general fearfulness. Increased dog-directed aggression was associated with a more positive judgement bias, while higher levels of fear predicted better discrimination of stimuli.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I have been fortunate in the support I have received over the course of my research, which enabled subject recruitment and data collection. I would like to thank Ahimsa Dog Training, LLC for donating the use of their dog training facility for research purposes, the private donors who contributed to an Experiment.com fundraising effort for research supplies, and the Animal Behavior Society and Association of Professional Dog Trainers Foundation for their research grants.

I am grateful to my committee for their guidance throughout this process. In particular, Dr. James Ha provided early support and inspiration that led me to pursue a graduate degree, and Dr. Robin Foster has offered tremendous academic and logistical support during the development and execution of this research. I am also indebted to the numerous dogs (and their owners) that helped with pilot data, and with training research assistants, especially Mavis and Potato Chip.

I would like to thank my friends and family members for their patience and encouragement throughout this process. My partner, Eduardo, has been a welcome source of humor, stability, and adventures outside of academia. Finally, I am eternally grateful to Fiona, for her moral support and for the incredibly tolerant, friendly nature that made her a perfect stimulus dog for my research.

DEDICATION

For Fiona, the most patient helper dog there ever was,

though she would rather have been rolling in dead worms.

Chapter 1. APPLICATIONS OF RESEARCH IN NONHUMAN ANIMAL PERSONALITY

Carolyn J. Loyer, James C. Ha

Abstract

Nonhuman animal (NHA) personality research has increased dramatically in recent years and encompasses a vast range of species, traits examined, and methods of measurement. To make use of this research, it is necessary to understand how various approaches interact with and complement each other, as well as examine the challenges posed for researchers in this field. In this chapter we describe a variety of ways NHA research has been or has the potential to be applied to current problems faced by both NHAs and humans. The potential applications for this field of research are as far-reaching as the approaches used to examine NHA personality.

1.1 INTRODUCTION

At different points in history it has been declared that only humans have emotions, use tools, or have a concept of "self", only to have researchers discover at least one exception to each of these rules shortly after they began looking for them. Rats *(Rattus norvegicus)* and dogs *(Canis lupus familiaris)* have been found to have specific ways of breathing or vocalizing that are akin to the human expression of laughter (Panksepp, 2000, 2007; Panksepp & Burgdorf, 2003; Simonet, Versteeg, & Storie, 2005). Charles Darwin's book "The Expression of Emotions in Man and Animals" documents multiple comparisons between human and nonhuman emotional responses. New Caledonian crows *(Corvus moneduloides)*, for instance, do not only use tools, they *create* them, modifying the sticks they use as "grub extractors" to improve their usefulness (Hunt & Gray, 2004; Weir, Chappell, & Kacelnik, 2002.) . Recognition of "self" in the ubiquitous mirror task has been observed in species ranging from chimpanzees (*Pan troglodytes)* (Gallup, 1970) to elephants *(Elephas maximus)* (Plotnik, de Waal, & Reiss, 2006). As each of these distinctions has been refuted, the line between humans and the rest of the animal kingdom has been blurred further.

Although nonhuman animals (henceforth: NHA) are used as models for human-centric psychological and medical treatment problems ranging from PTSD, parental attachment, and autism to cancer treatment and medical emergencies, until the last few decades it was generally considered overly anthropomorphic to focus on NHA personality. Even fundamental statistical designs focused on group means rather than individual differences. However, research in the field of individual differences in NHAs has recently (in the last twenty years or so) transitioned from taboo to popularity. With this recent surge in NHA personality research, it has become clear that nonhuman animals do, in fact, display consistent individual differences in their patterns of behavior across contexts and time. To put it bluntly, they have "personality". Other chapters of this book outline the analogues between human and NHA personality and give the reader a taste of the "basic science" discoveries that NHA personality research has uncovered in terms of the evolutionary and biological underpinnings of individual differences. The goal of this chapter is to explain *why* it's important that NHAs have personality, and to outline some of the practical applications of research in this area.

Beyond the interest in individual differences for its own sake, a majority of behavioral research is focused on predicting behavior. Most research looks at average increases or decreases in behavior across a sample when the independent variable changes, e.g., average increase in singing by song sparrows *(Melospiza melodia)* as a function of the season, presence of potential mates, or parasite load. In cases like these, individual differences are seen as statistical "noise" obscuring the population-level effects in which the researcher is actually interested (a problem also noted by Réale, Reader, Sol, Mcdougall, & Dingemanse, 2007). Nonetheless, a better understanding of that noise allows a researcher to better control for it, improving the validity and generalizability of the population-level results, and decreasing the sample sizes required for statistical significance, which is of both practical and ethical value.

1.2 WITHIN-SPECIES APPLICATIONS

Although a large portion of research conducted on NHAs is for "basic science," and NHA personality certainly has plenty to offer in terms of basic discoveries, improved understanding of individual differences has enormous potential to benefit not just the species being studied, but the subjects themselves. For NHAs in the care of humans, we have a responsibility and often a vested interest in providing living conditions that maximize their normal range of behavior and wellbeing both physically and emotionally. From an ethical standpoint, given that we choose to keep NHAs in captivity for entertainment, companionship, for research, to assist in our daily lives, and to provide food, we have an equal duty to ameliorate the potential negative effects of that captivity. Standards of animal care in private homes, in zoos and aquariums, and in research laboratories have improved drastically in the last century and a half (Gauthier & Griffin, 2005). We've gone from providing the bare minimum in terms of shelter and food to designing elaborate, naturalistic enclosures, providing opportunities for NHAs to use their basic hunting and foraging skills, and arranging social housing when it is species-appropriate. These changes reflect a major shift not only in our concern for NHAs' physical health, but also towards our responsibility to maximize their psychological wellbeing. Scientists have repeatedly demonstrated that providing these types of environmental and social enrichment can increase activity and may improve medical and/or reproductive outcomes (Johansson & Ohlsson, 1996; Meagher et al., 2014; Passineau, Green, & Dietrich, 2001). However, NHA caretakers have long observed that environmental enrichment or deprivation do not affect all NHAs' behavior equally.

1.2.1 Environment/Personality interactions

In collaboration with the Woodland Park Zoo, we investigated behavioral differences in three elephants(two were Asian elephants, *Elephas maxima,* and one was African, Loxodonta) that had been sharing the same space for the majority of a thirty-plus year span. Bamboo, Chai, and Watoto (the African elephant) shared their enclosure at the Woodland Park Zoo, but though all three had similar high standards of care and enrichment opportunities, their keepers noticed behavior patterns distinct enough from each other that the zoo wanted to better understand. Specifically, Chai had developed a very distinctive rocking stereotypy that both intrigued and concerned zoo visitors. Bamboo and Watoto, on the other hand, showed very little rocking behavior. The zoo had been incorporating more and more enrichment in the form of toys, food puzzles, naturalistic browse, and training with the keepers, and wanted to determine whether those efforts had had an impact on the stereotypical behavior of any or all of the elephants. They recruited researchers at the University of Washington to oversee data collection and analysis of the elephants' behavior throughout the day. Using focal scan sampling, researchers catalogued what the elephants did at various times of day for four years. The results were consistent with staff and volunteer observations: Chai spent a greater proportion of her time engaged in stereotypic behavior than did Bamboo or Watoto, and the *type* of stereotypy she engaged in was different than Bamboo and Watoto.

Examining the elephants' range of behaviors provided additional information: the activity that Chai's rocking behavior replaced. During the times of day when Chai was most likely to be seen rocking, Bamboo and Watoto dramatically increased their independent foraging activity. The perimeter of the elephant enclosure was intentionally planted with live, edible foliage, which the elephants regularly snagged and devoured as part of their enrichment program. During the hour or so leading up to a regular feeding, Bamboo and Watoto would often snack on the plants around the perimeter, while Chai was more likely to spend that time in anticipatory rocking. This information suggested that providing Chai with more foraging opportunities, or "training" her to exploit existing opportunities as the other elephants were already doing, might be an effective intervention in reducing her rocking behavior (Loyer, Ha, Fernandez, & Hawkes, 2013). The results of this study were promising: across the observation years (and correlated with increased enrichment opportunities) there was a decrease in stereotypy across all of the elephants. This study supports the thinking that enrichment, or added environmental complexity, can reduce the frequency and duration of stereotypic behaviors, and that NHAs have consistent individual differences in their responses to environmental enrichment, suggesting that enrichment may need to be targeted to an NHA's personality to have maximum impact.

1.2.2 Effect of environmental enrichment

One simple explanation for differences in the efficacy of environmental enrichment is that some NHAs may simply be too afraid to interact with the enrichment provided. This would explain why, when new structures and methods of food delivery are available, some individuals show a substantial increase in activity while others show no change. Walker and Mason (2012) tested this hypothesis by first assessing the degree of neophobia (fear of novel objects or situations) in female mice *(Mus musculus)* by placing novel objects into their home cages and calculating how long it took them to make contact with the object. Then they gave the mice free access to an enriched cage environment that included a running wheel, objects to chew on, nesting materials, complex surfaces for the mice to climb on and around, and a variety of manipulable toys. The researchers measured the use of the enriched environment by the amount of food consumed in the enriched cage compared to their standard laboratory cages, and by the quantity of two "consumable" forms of enrichment (a cardboard planter pot, which the mice shredded, and a length of string that they could pull into the cage to chew or use for nesting materials). They found that mice with higher levels of neophobia (those who took longer to interact with a novel object in their cage) tended to eat less food in the enriched cage and to consume less of the available enrichment. These results support the hypothesis that enrichment items may actually be frightening for individuals that are more fearful. They also beg the question of whether neophobia might be indicative of a general personality trait of fearfulness. In such a case, we'd predict that neophobic individuals would show extra caution across multiple scenarios, for example a greater response to an environmental stressor or increased vigilance under threat of predation. Alternatively, if no relationship (or even an opposite relationship) was found between neophobia and behavior in other contexts, it opens an entirely new line of inquiry as we try to explain why (evolutionarily, biologically, or ontogenetically) an individual might be fearful in one context and not in others. Occasionally studies like this deliver reasonably straightforward results, but it is far more common for researchers to find a complex relationship between behavior and environment.

The University of California at Davis keeps a colony of orange-winged Amazon parrots *(Amazona amazonica)* for research and has conducted multiple behavioral studies with them that examine the relationship between environmental enrichment and personality. Previous research at the university demonstrated that parrots reared in enriched environments tend to be less neophobic than those reared in barren or completely stable environments (Cussen & Mench, 2015). To test whether reduction in neophobia is due to the enrichment items themselves or the consistent exposure to novelty, Fox and Millam (2007) compared neophobia (measured as the latency to eat something tasty from a feeder with a new object hanging nearby) before, during, and after orange-winged Amazon parrots were exposed to eleven weeks of either stable environmental enrichment objects (the low novelty condition), or rotating enrichment items (the high novelty condition). On a group level, the high novelty treatment had a greater impact on neophobia than the low novelty treatment. However, for individuals that were extremely neophobic prior to the treatment, their neophobia actually *increased* in the high novelty condition compared to those in the low novelty condition. These results indicate that *exposure to novelty is only an effective treatment for neophobia if that exposure is not too severe for the individual being exposed.* The implications for human behavioral treatments of fear, anxiety, and neophobia-related conditions are immediately obvious.

1.2.3 Response to barren environments

The previous two studies examined how enriched environments can improve potentially negative *personality traits and how personality traits affect NHA responses to enrichment.* At the opposite end of the spectrum, scientists have studied how NHAs vary in their response to barren environments based on personality. One common metric of welfare used in zoos and laboratories is the amount of time animals spend engaged in repetitive behaviors called stereotypies. These are behaviors like the rocking we previously described in Chai the elephant, or the famous pacing of caged predators(Clubb & Mason, 2003; Mason, 2010; Clubb & Vickery, 2006). There are quantifiable differences between conspecific individuals housed in similar conditions or even in the same enclosures when it comes to the degree and type of stereotypic behavior they exhibit (Cussen & Mench, 2015; Loyer, Ha, Fernandez, & Hawkes, 2013), health outcomes (Cavigelli, 2005), and social relationships (Weinstein & Capitanio, 2008).

Cussen and Mench (2015) looked at how feather plucking (a common problem behavior in captive psittacines) and locomotor stereotypies were affected by 20 weeks of barren housing in orange-winged Amazon parrots at the UC Davis colony that had previously been rated on what the authors labeled "extraversion" and "neuroticism." These ratings were obtained by having two people familiar with all of the birds assign a value from 0 to 7 based on how well an adjective described their personality. Birds that were considered high on the factor they labeled neuroticism were aggressive, eccentric, excitable, fearful, inflexible, high-strung, shy, and were not affiliative, tame or confident. Birds that were considered highly extraverted were active, bold, impulsive, intelligent, cooperative, persistent, and were not cautious or lazy. Prior to enrichment deprivation, the researchers collected behavioral data which was repeated at the end of deprivation and after they had reintroduced enrichment.

There were some fairly universal changes in behavior: when deprived of enrichment, birds spent more time preening (cleaning and straightening out their feathers with their beaks), more time engaged in locomotor stereotypies, and more time generally active during the barren housing treatment than in either the baseline or the re-enriched conditions. Additionally, the birds' feather condition deteriorated significantly during the barren condition, indicating feather plucking and over-preening (and although they returned to baseline levels of preening after enrichment was reintroduced, the birds' feather condition did not improve even after the enrichment was returned). While the changes described above were found in nearly all birds, there were also significant differences that were predicted by individual parrots' personality ratings. Birds who were rated as more "neurotic" had poorer feather condition, even though they didn't spend any more time preening than their less neurotic counterparts. Birds who were higher on extraversion showed a smaller increase in stereotypy after the barren environment and after re-enrichment: The deprivation had a smaller impact on their stereotypy than birds that were less extraverted. This difference in sensitivity to stressful environments based on personality has been found in multiple species, and although the welfare implications should be sufficient motivation to continue research in this line, research performed on pigs at an experimental farm at a university in the Netherlands offers additional arguments for understanding how personality and environment interact.

Bolhuis, Schouten, Schrama, and Wiegant (2006) investigated how different rearing environments as well as housing environment later in life impacted domestic pigs *(Sus scrofa domesticus)* on the experimental farm at Wageningen University. They looked at behavior as well as health to determine how personality or coping style interacted with the environment. Pigs' personalities were assessed using the "Backtest" in the first few weeks of life (see also Horback, this volume). At 10 days old, experimenters held the piglets on their backs for sixty seconds and counted the number of times they struggled during the restraint. The test was repeated at 17 days old, and the escape attempts were added across the two tests in an effort to get a more global picture of the pigs' behavior. They categorized piglets as high resisting (HR) or low resisting (LR), and then studied the responses of both groups of pigs to two different housing environments: barren floors versus straw bedding. The researchers manipulated whether pigs had access to straw during rearing (from weaning until 10 weeks old), and later during the "finishing phase" (the period between 10 weeks old and 22 weeks old, when they are slaughtered). One quarter of the pigs spent the entire time with straw bedding available, another quarter spent the entire time on barren substrate, and the remaining pigs started with either barren floors or straw and were switched at 10 weeks onto the opposite flooring. During the study, they looked at the pigs' behavior, overall health and weight gain, and found complicated interactions between rearing environment, later housing environment, and coping style. HR pigs tended to be more aggressive than LR pigs, and the LR pigs tended to be more sensitive to the environment and its changes: while HR pigs in barren environments showed no difference in the amount of time they spent chewing/biting at their penmates based on their rearing environment, LR pigs reared in barren environments spent more time chewing on their penmates, especially if they transitioned to enriched housing during their finishing phase. LR pigs also played more on straw bedding than on barren floors, particularly if they were also reared on straw.

These findings on behavioral differences have welfare implications for livestock raised for human consumption and should be sufficient on their own to warrant further investigation. However, the researchers also found that personality and environment interacted when it came to health and weight in the pigs, which has immediate practical implications for farming protocols. There were significant differences in the average daily weight gain in HR pigs raised in enriched environments: If they stayed in an enriched environment, they gained about 10% more weight during the finishing phase than if they were switched to barren floors. They also found that LR pigs had a significant decrease in the occurrence of gastric lesions when they spent the second half of their lives on straw bedding. Although these results present a substantial argument in favor of environmental enrichment, they also highlight the fact that individual differences in behavior mean different ways of expressing stress, different degrees of response to particular stressors, and a very clear need for better understanding of each of these facets of NHA personality. They also suggest that selective breeding for personality or behavior in addition to physical attributes has the potential to improve productivity on farms with minimal cost. While this research was conducted with pigs, other studies examining livestock coping styles/behavioral syndromes/personalities have found similar variation in behavior that could inform husbandry and breeding practices to both improve animal welfare and increase efficiency and productivity on farms (Lansade et al., 2014; Müller & Schrader, 2005). Because research in NHA personality is relatively new, scientists have much more work to do in testing the generalizability of findings from one species to the next. Still, there is sufficient evidence to indicate that some basic personality traits are comparable even across taxa (Gosling, 2001; Gosling & John, 1999).

1.3 ACROSS-SPECIES APPLICATIONS

An example of research in one species having implications for many others comes from a study by Ferland et al. (2014). Ferland and her colleagues investigated the complex relationship between an individual's sensitivity to novelty, and the effect of repeated exposure to that novelty on impulsive behavior in Long-Evans rats *(Rattus norvegicus)*. In rats, one way to measure motor impulsivity is by looking at their ability to essentially "wait their turn" in a task where they have to choose the correct hole out of five options in order to earn a sugar pellet. Between trials, the rats are required to wait a whopping five seconds before a light indicates which hole is the correct choice for the next trial. Poking their nose into any of the holes before the 5 seconds is up is an example of motor impulsivity, and it turns out that in *some* rats, you can drastically increase the number of premature responses by repeatedly exposing them to a stimulating, enriched environment. Researchers assigned half of their rats to the enriched environment condition, and for 16 days would place them in a large, enriched cage for an hour prior to testing while the other half remained in their normal housing during that hour. Some rats began nose-poking prematurely far more often at the end of the stimulating environment condition than they had during baseline testing, so the researchers repeated the experiment with a second cohort of rats. This time, at the start of the experiment they used infrared beams to measure the rats' amount of locomotor activity in a novel environment for 60 minutes prior to beginning the task training. Because not all rats in the enriched environment condition had changed their proportion of premature responses, the experimenters were trying to identify any pre-existing behavioral differences between rats that were impacted by the environmental enrichment and those who were not. They found a significant difference in activity between rats whose premature responses increased drastically in the stimulating environment condition and those who did not: Rats that were more sensitive to the extra stimulation were also more active in the first 40 minutes in a novel environment, but by the end of their 60 minutes, they were no more active than the less sensitive rats. Thus, a consistent personality trait appeared to govern both their exploratory behavior in novel environments, and their impulsivity in response to environmental enrichment.

The implications of this research (and other experiments like it) are numerous. It demonstrates the complicated interaction between NHA personality and environment (hyperactivity may be related to an increased sensitivity to novel stimuli, which may result in an increase in impulsive behavior, which, in humans, can be considered part of the personality itself!). If this experiment could be applied to species used for assisting humans, it may help inform our decisions about how to more effectively train animals, like dogs, or which individuals to use for particular tasks.

1.4 SELECTION AND EVALUATION OF ANIMALS FOR SPECIFIC TASKS

1.4.1 Working animals

Consider, for a moment, the behavioral requirements of a guide dog tasked with helping a blind person maneuver around a busy metropolitan area. Impulsivity is an enormously risky trait when it might result in a handler being yanked into traffic, and it is one reason that dogs "fail out" of the training (Burrows, Adams, & Millman, 2008). If some dogs, like rats, are likely to become more impulsive after prolonged exposure to a highly stimulating environment (such as a big city), and if those dogs can be identified before extensive time and money have been spent preparing them to be guide dogs, guide dog organizations (and the individuals they benefit) could save a substantial investment that might have otherwise been wasted. A few studies have examined different behavior and personality assessment techniques to determine their efficacy in predicting success of canine candidates, with varied outcomes.

The earlier canine candidates can be identified, and the sooner unsuitable dogs can be eliminated from the program, the more efficiently working dog programs can function. Breeding and training working dogs is enormously expensive, ranging from around $19,000 for a police dog (http://www.nationalpolicedogfoundation.org) to $42,000 for a seeing-eye dog (http://www.guidedogsofamerica.org). In addition to the financial cost, training for these programs constitutes a serious investment of time (an estimated 18-24 months for service or assistance dogs; http://servicedogcentral.org/; http://www.assistancedogsinternational.org). Therefore, the cost of training an individual dog that is later ejected from the program due to unsuitable behavior is immense. On the other hand, if the methods used for early rejection are unreliable, excellent candidates will be removed from the program before they have a chance to demonstrate their suitability. Goddard and Beilharz (1986) tested Labrador puppies that were purpose-bred to be guide dogs to determine how early important behavioral and personality differences could be identified to help with selection of both breeding stock and dogs that would be successful in the training program. They found that consistent differences in fearfulness (one of the most important factors in a dog's success as a guide dog) could be detected as early as eight weeks of age, but that tests to predict adult fearfulness were more accurate as the dog's age increased. Based on their research, Goddard and Beilharz suggest that genetic selection for fear in breeding stock is most accurate in adulthood, which is crucial information for working dog breeding programs. Many programs neuter males that will be sent through training when they are adolescents, which means that dogs that may have potentially valuable genetic contributions are eliminated from the breeding pool before their potential can be adequately assessed. Wilsson and Sundgren (1998) were interested in this problem, and tested whether tests conducted on eight week old puppies could adequately predict adult behavior. They found that the "… correspondence of puppy test results to performance at adult age was negligible" (p. 1).

This type of NHA personality research, aimed at determining which personality traits stabilize at different stages ontogenetically, can have immediate practical impacts in addition to feeding back into broader theories of personality development. Additionally, research that helps identify the most accurate and efficient method(s) and age(s) for assessing NHA personality can improve our selection process for purposeful breeding in species we live and work with. These results should be beneficial not only for working canines, but for livestock and companion animals as well.

In an effort to produce results comparable across species, multiple NHA studies have imitated methods used in human research by having observers familiar with the NHAs being assessed rate the degree to which adjectives apply to the individuals (Gosling, 1998) or by asking about individual responses to specific events (Hsu & Serpell, 2003). Unfortunately, these methods depend on the availability of familiar observers (which is sometimes impossible), and introduce a risk of subjectivity and anthropomorphization of NHA behavior. Thus, behavioral coding is often used instead, although this brings its own set of limitations, one of the most serious being the heavy time and resource investment required. To run a single individual through a battery of tasks designed to assess multiple facets of personality can require hours, even if these tasks are conducted in a single environment. Many researchers avoid this by using extremely simple assessment tools that measure behavior differences on one or perhaps two axes (Amy, van Oers, & Naguib, 2012; Korsten et al., 2010; Sinn, Apiolaza, & Moltschaniwskyj, 2006). This simplification of personality to one or two dimensions is useful for establishing that individual differences exist, but does not feed back into broader theory about how personality and environment interact, as it tests extremely limited slices of each. In the same vein of critique, behavioral assays are often implemented in a laboratory setting, which calls into question the generalizability to related behaviors in the wild or in real-life circumstances, or worse, whether the tasks themselves are sufficiently biologically relevant to allow interpretation of the results.

The degree to which the limitations of these techniques hobble their practicality depends on their purpose. If the goal is to breed more docile animals to improve the ease of handling them in captivity, it is probably sufficient to perform simple assays by well-trained observers who need not know the animals individually. Dmitri Belyaev's famous work with silver foxes *(Vulpes vulpes)* in Russia may be the best proof of this. Using a simple two-part assay, Belyaev and his colleagues selected individuals that showed minimal signs of fear or aggression towards humans and bred them. Within 10 generations, almost twenty percent of the selectively bred population were what Belyaev labeled "domestic elite," meaning that, when interacting with humans, they would whimper, sniff the experimenters' hands, and lick them much like domestic dogs do (Trut, 1999). In this case, the researchers were interested in domestication and in improving ease of handling, while looking at the behavioral, genetic, and physiological side effects of this artificial selection. This experiment has provided important information for basic science about the genetics of behavior, and offers a starting point for those interested in selecting animals for captivity or livestock based on improving ease of handling. However, the subjects of this research were bred and raised specifically for this project, and therefore had fulfilled their duties simply by demonstrating more affiliation with humans. In more applied settings, the personality or behavioral tasks required of NHAs are often far more complex, and the humanitarian and economic costs of improper breeding and selection of may be far more severe. As a result, more complicated personality or behavior measures are often required. Determining the best method of measurement for each purpose is one subset of applied NHA personality, and again, the domestic dog provides an opportune species for investigation.

In an experiment that compares multiple ways of evaluating working dog behavior, Rooney, Gaines, Bradshaw, and Penman (2007) compared the ratings of dogs' handlers/trainers, experienced trainers who were unfamiliar with the subject dogs, and scientists, as well as objective measures obtained during an assessment completed at the end of a training program at the Defence Animal Centre in the UK. The trainers rated the dogs throughout their ten-week training program. At the end of this program, the dogs were run through a series of search tasks which were video recorded and rated independently by scientists and experienced military dog trainers who were unfamiliar with the subjects. These independent observers rated the dogs on similar (but not identical) characteristics to those assessed by their trainers. Finally, the recordings of the search tasks were analyzed for behavior measures thought to be important for a dog's success as a working detection dog. The outcomes of each of these methods of assessing the dogs' behavior were compared. Ideally, if each method is equally valid, the agreement between the methods would be high, which is what they found: The trainers who had worked with the dogs for ten weeks and knew them rated their abilities similarly, the independent scientists and unfamiliar trainers who rated the dogs based solely on the video recordings rated the dogs very similarly to one another, and overall, the trainers' ratings and independent scientist and trainer ratings for the dogs were correlated with each other and with the objective measures. This type of research is crucial to our ability to identify the best way to measure NHA's personality and working potential. If it's equally valid to use people familiar with an individual to rate their overall personality or behavior, versus people familiar with the species and the tasks required for the circumstances, or to use ethological measures that can be assessed by someone who need not be familiar with the species or the tasks as long as they can be trained to measure the crucial criteria, it allows us to a) pick the best measurement method based on time, cost, and personnel available, and b) more easily compare outcomes across studies that utilize different methods to assess personality and behavior.

1.4.2 Selection of companion animals

In addition to the human-centric importance of using effective behavior and personality assessments for working dogs, most NHAs kept as pets stand to benefit from improving the tests used to assess their performance and personality. Behavior assessments are used in animal shelters and at rescues to determine the right type of home environment for the NHAs, as well as to identify NHAs that present a public safety threat. A test that is supposed to measure a dog's bite risk necessarily has a major impact on the dog's life: with limited space and resources in shelters, dogs are often euthanized based on behavior rather than medical necessity. In these cases, knowing which (if any) assay can accurately predict bite risk (the chances that they bite a human or another NHA) could potentially save NHA lives while improving safety for the general public.

Currently, most behavioral assessments used by animal shelters, breeders, or rescue organizations have not been scientifically validated. Furthermore, they tend to be blunt instruments aimed at predicting bite risk (Taylor & Mills, 2006), as opposed to understanding the nuances of behavior that may promote or degrade their relationship with prospective adopters. Tests of common assessments show a limited ability to differentiate dogs with a known history of aggression from those who have no aggressive history (Bennett, Litster, Weng, Walker, & Luescher, 2012; Bollen & Horowitz, 2008; Paroz, Gebhardt-Henrich, & Steiger, 2008), calling into question their utility as a tool for determining adoptability and placing dogs with appropriate families. Given the stakes of these assessments, this is an area of applied research in NHA personality that is in need of far more attention.

1.5 Human applications

The studies outlined previously have covered a range of applied research in NHA personality that is useful for the sake of understanding NHA personality and behavior in itself. In some cases, it is useful for ethical reasons and can benefit NHAs themselves. In others, it is primarily beneficial to the humans living and working with these NHAs, and in still other circumstances, it is beneficial to NHAs and humans alike. Besides the utility of understanding NHA personality for its own sake, NHAs have been used as models for human behavior and medicine essentially since the beginning of scientific enquiry.

We use NHAs as models for humans when it is impractical or unethical to conduct experiments on humans. It is not uncommon to deliberately breed or genetically engineer animals to specifically exhibit a medical or behavioral disorder, in order to identify factors that might affect that disorder's presentation, or to test treatments. Some strains of laboratory rats and mice are more likely to show fearful behavior (Higley, Hasert, Suomi, & Linnoila, 1991), mice have been bred for their propensity to display stereotypies (Muehlmann et al., 2012) so that they can be used as models for autism treatment, and we examine the relationship between exploratory behavior or neophilia and genetics in species ranging from vervet monkeys *(Chlorocebus pygerythrus)* (Bailey, Breidenthal, Jorgensen, McCracken, & Fairbanks, 2007) to great tits *(Parus major)* (Fidler et al., 2007; Korsten et al., 2010). With NHAs, the possibility of selective breeding combined with decreased time between generations allows researchers to get answers faster than they would with human subjects. The availability of species with social lives that are similarly complex to those of humans allows for research on interactions between different personality types, and NHA models are the only experimental option when testing the effect of major stressors on personality development. In humans, personality has been linked to differential outcomes in a variety of areas, including substance use and abuse (Cloninger, Sigvardsson, & Bohman, 1988; Kotov, Gamez, Schmidt, & Watson, 2010), academic performance (Chamorro-Premuzic & Furnham, 2003; Komarraju, Karau, & Schmeck, 2009), and relationship outcomes and satisfaction (Shiota & Levenson, 2007; Vohs, Finkenauer, & Baumeister, 2011). Thus, if we are to use NHAs to understand human behavior and medical issues, it is critical that we at least attempt to account for the individual differences that might affect the conclusions we are attempting to make.

1.6 Conclusion

Applications of NHA personality research are numerous, and span multiple fields from conservation (see Smith, this volume; Watters, Powell & Bermner-Harrison, this volume), captive animal welfare (see Horback, this volume; Lilly & Yeater, this volume), and the safety of the general public, to medicine and evolutionary biology (Brommer et al., this volume; Van Oers

& Laine, this volume), and the list goes on. Given the vast number of areas and species in which this research can be applied, perhaps it is no surprise that we have yet to come up with a universal method of assessment, and indeed, that universality is probably not only difficult (if not impossible) to achieve, but inadvisable. Depending on the specific aspects of an individual's personality or behavior that need to be understood, assays must be tailored to be appropriately broad or specific to the task. Perhaps one way to consider this challenge is by comparing NHA personality tests to aptitude tests, much like many high school students take to determine "what they should be when they grow up." The goal of these tests is to ensure that the right individuals end up in the right vocations, but there are two ways one might get to the same point: 1) Develop a general assessment that examines behavior in a variety of contexts and try to use this assessment to direct individuals to the best task, or 2) develop an assessment that is geared at differentiating candidates' aptitude at a specific task with a simple "yes or no" answer at the end. Each approach has its advantages and disadvantages. Whereas a broader assessment is better able to examine how facets of behavior and personality interact, it may necessarily sacrifice the ability to predict behavior in specific circumstances as the result of its broader perspective. Alternatively, the simpler assessment that examines small pieces of behavior under limited conditions may be more accurate at predicting behavior in similar circumstances in the future, but may, in isolation, tell us less about the broader construct of personality. However, these slices of information, when combined, can help us piece together a better understanding of personality (or individual differences, or coping styles, or behavioral phenotype, or whatever one chooses to call it), in both human and nonhuman animals. NHA personality research has enormous potential to improve the lives of all animals, human and otherwise.

References

Amy, M., van Oers, K., & Naguib, M. (2012). Worms under cover: Relationships between performance in learning tasks and personality in great tits (Parus major). *Animal Cognition*, *15*(5), 763–770. http://doi.org/10.1007/s10071-012-0500-3

Bailey, J. N., Breidenthal, S. E., Jorgensen, M. J., McCracken, J. T., & Fairbanks, L. a. (2007). The association of DRD4 and novelty seeking is found in a nonhuman primate model. *Psychiatric Genetics*, *17*(1), 23–27. http://doi.org/10.1097/YPG.0b013e32801140f2

Bennett, S. L., Litster, A., Weng, H. Y., Walker, S. L., & Luescher, A. U. (2012). Investigating behavior assessment instruments to predict aggression in dogs. *Applied Animal Behaviour Science*, *141*(3–4), 139–148. http://doi.org/10.1016/j.applanim.2012.08.005

Bolhuis, J. E., Schouten, W. G. P., Schrama, J. W., & Wiegant, V. M. (2006). Effects of rearing and housing environment on behaviour and performance of pigs with different coping characteristics. *Applied Animal Behaviour Science*, *101*(1–2), 68–85. http://doi.org/10.1016/j.applanim.2006.01.001

Bollen, K. S., & Horowitz, J. (2008). Behavioral evaluation and demographic information in the assessment of aggressiveness in shelter dogs. *Applied Animal Behaviour Science*, *112*(1–2), 120–135. http://doi.org/10.1016/j.applanim.2007.07.007

Burrows, K. E., Adams, C. L., & Millman, S. T. (2008). Factors affecting behavior and welfare of service dogs for children with autism spectrum disorder. *Journal of Applied Animal Welfare Science : JAAWS*, *11*(1), 42–62. http://doi.org/10.1080/10888700701555550

Cavigelli, S. (2005). Animal Personality and Health. *Behaviour*, *142*(9/10), 1223–1244. http://doi.org/10.1163/156853905774539355

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality*, *37*(4), 319–338. http://doi.org/10.1016/S0092-6566(02)00578-0

Cloninger, C. R., Sigvardsson, S., & Bohman, M. (1988). in Young Adults, *12*(4), 494–505.

Clubb, R., & Mason, G. (2003). Captivity effects on wide-ranging carnivores. *Nature*, *425*(October), 473–474.

Cussen, V. A., & Mench, J. A. (2015). The Relationship between Personality Dimensions and Resiliency to Environmental Stress in Orange-Winged Amazon Parrots (Amazona amazonica), as Indicated by the Development of Abnormal Behaviors. *PLoS ONE*, *10*(6), e0126170. http://doi.org/10.1371/ journal.pone.0126170

Ferland, J. M. N., Zeeb, F. D., Yu, K., Kaur, S., Taves, M. D., & Winstanley, C. A. (2014). Greater sensitivity to novelty in rats is associated with increased motor impulsivity following repeated exposure to a stimulating environment: Implications for the etiology of impulse control deficits. *European Journal of Neuroscience*, *40*(12), 3746–3756. http://doi.org/10.1111/ejn.12748

Fidler, A. E., van Oers, K., Drent, P. J., Kuhn, S., Mueller, J. C., & Kempenaers, B. (2007). Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proceedings. Biological Sciences / The Royal Society*, *274*(1619), 1685–91. http://doi.org/10.1098/rspb.2007.0337

Fox, R., & Millam, J. (2007). Novelty and individual differences influence neophobia in orange-winged Amazon parrots (Amazona amazonica). *Applied Animal Behaviour Science*, *104*(1–2), 107–115. http://doi.org/10.1016/j.applanim.2006.04.033

Gauthier, C., & Griffin, G. (2005). Landmarks in animal-based research and key moral. *Revue Scientifique et Technique (International Office of Epizoonotics)*, (September 2005).

Goddard, M. E., & Beilharz, R. G. (1986). Early prediction of adult behaviour in potential guide dogs. *Applied Animal Behaviour Science*, *15*(3), 247–260. http://doi.org/10.1016/0168-1591(86)90095-X

Gosling, S. D. (1998). Personality dimensions in spotted hyenas (Crocuta crocuta). *Journal of Comparative Psychology*, *112*(2), 107–118. http://doi.org/10.1037/0735-7036.112.2.107

Gosling, S. D. (2001). From mice to men: what can we learn about personality from animal research? *Psychological Bulletin*, *127*(1), 45–86. http://doi.org/10.1037/0033-2909.127.1.45

Gosling, S. D., & John, O. P. (1999). Personality Dimensions in Nonhuman Animals: A Cross-Species Review. *Current Directions in Psychological Science*, *8*(3), 69–75. http://doi.org/10.1111/1467-8721.00017

Higley, J. D., Hasert, M. F., Suomi, S. J., & Linnoila, M. (1991). Nonhuman primate model of alcohol abuse: effects of early experience, personality, and stress on alcohol consumption. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(16), 7261–7265. http://doi.org/10.2307/2357662

Hsu, Y., & 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*(9), 1293–1300. http://doi.org/10.2460/javma.2003.223.1293

Hunt, G. R., & Gray, R. D. (2004). Direct observations of pandanus-tool manufacture and use by a New Caledonian crow (Corvus moneduloides), 114–120. http://doi.org/10.1007/s10071-003-0200-0

Johansson, B. B., & Ohlsson, A. L. (1996). Environment, social interaction, and physical activity as determinants of functional outcome after cerebral infarction in the rat. *Exp Neurol*, *139*(2), 322–327. http://doi.org/10.1006/exnr.1996.0106

Komarraju, M., Karau, S. J., & Schmeck, R. R. (2009). Role of the Big Five personality traits in predicting college students' academic motivation and achievement. *Learning and Individual Differences*, *19*(1), 47–52. http://doi.org/10.1016/j.lindif.2008.07.001

Korsten, P., Mueller, J. C., Hermannstädter, C., Bouwman, K. M., Dingemanse, N. J., Drent, P. J., … Kempenaers, B. (2010). Association between DRD4 gene polymorphism and personality variation in great tits: a test across four wild populations. *Molecular Ecology*, *19*(4), 832–43. http://doi.org/10.1111/j.1365-294X.2009.04518.x

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: a meta-analysis. *Psychol Bull*, *136*(5), 768–821. http://doi.org/10.1037/a0020327

Lansade, L., Valenchon, M., Foury, A., Neveux, C., Cole, S. W., Layé, S., … Moisan. (2014). Behavioral and transcriptomic fingerprints of an enriched environment in horses (Equus caballus). *PLoS ONE*, *9*(12), 1–20. http://doi.org/10.1371/journal.pone.0114384

Mason, G. J. (2010). Species differences in responses to captivity: Stress, welfare and the comparative method. *Trends in Ecology and Evolution*, *25*(12), 713–721. http://doi.org/10.1016/j.tree.2010.08.011

Clubb, R., & Vickery, S. (2006). Locomotory stereotypies in carnivores: Does pacing stem from hunting, ranging, or frustrated escape? In Mason, G., & Rushen, J. (Eds.) *Stereotypic Animal Behaviour:Fundamentals and Applications to Welfare Second Edition*(pp. 58-85). Cambridge, MA: CABI.

Meagher, R. K., Dallaire, J. A., Campbell, D. L. M., Ross, M., Møller, S. H., Hansen, S. W., … Mason, G. J. (2014). Benefits of a ball and chain: Simple environmental enrichments improve welfare and reproductive success in farmed american mink (Neovison vison). *PLoS*

ONE, *9*(11). http://doi.org/10.1371/journal.pone.0110589
Muehlmann, a M., Edington, G., Mihalik, a C., Buchwald, Z., Koppuzha, D., Korah, M., & Lewis, M. H. (2012). Further characterization of repetitive behavior in C58 mice: developmental trajectory and effects of environmental enrichment. *Behavioural Brain Research*, *235*(2), 143–9. http://doi.org/10.1016/j.bbr.2012.07.041
Muller, R., & Schrader, L. (2005). Behavioural Consistency during Social Separation and Personality in Dairy Cows. *Behaviour*, *142*(9), 1289–1306.
Panksepp, J. (2000). The riddle of laughter: Neural and psychoevolutionary underpinnings of joy. *Current Directions in Psychological Science*, *9*(6), 183–186. http://doi.org/10.1111/1467-8721.00090
Panksepp, J. (2007). Neuroevolutionary sources of laughter and social joy: Modeling primal human laughter in laboratory rats. *Behavioural Brain Research*, *182*(2), 231–244. http://doi.org/10.1016/j.bbr.2007.02.015
Panksepp, J., & Burgdorf, J. (2003). "Laughing" rats and the evolutionary antecedents of human joy? *Physiology and Behavior*, *79*(3), 533–547. http://doi.org/10.1016/S0031-9384(03)00159-8
Paroz, C., Gebhardt-Henrich, S. G., & Steiger, A. (2008). Reliability and validity of behaviour tests in Hovawart dogs. *Applied Animal Behaviour Science*, *115*(1–2), 67–81. http://doi.org/10.1016/j.applanim.2008.04.010
Passineau, M. J., Green, E. J., & Dietrich, W. D. (2001). Therapeutic effects of environmental enrichment on cognitive function and tissue integrity following severe traumatic brain injury in rats. *Experimental Neurology*, *168*(2), 373–384. http://doi.org/10.1006/exnr.2000.7623
Plotnik, J. M., de Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(45), 17053–17057. http://doi.org/10.1073/pnas.0608062103
Réale, D., Reader, S. M., Sol, D., Mcdougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. http://doi.org/10.1111/j.1469-185X.2007.00010.x
Rooney, N. J., Gaines, S. A., Bradshaw, J. W. S., & Penman, S. (2007). Validation of a method for assessing the ability of trainee specialist search dogs. *Applied Animal Behaviour Science*, *103*(1–2), 90–104. http://doi.org/10.1016/j.applanim.2006.03.016
Shiota, M. N., & Levenson, R. W. (2007). Birds of a feather don't always fly farthest: similarity in Big Five personality predicts more negative marital satisfaction trajectories in long-term marriages. *Psychology and Aging*, *22*(4), 666–675. http://doi.org/10.1037/0882-7974.22.4.666
Sinn, D. L., Apiolaza, L. a, & Moltschaniwskyj, N. a. (2006). Heritability and fitness-related consequences of squid personality traits. *Journal of Evolutionary Biology*, *19*(5), 1437–47. http://doi.org/10.1111/j.1420-9101.2006.01136.x
Taylor, K. D., & Mills, D. S. (2006). The development and assessment of temperament tests for adult companion dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, *1*(3), 94–108. http://doi.org/10.1016/j.jveb.2006.09.002
Trut, L. N. (1999). Early canid domestication: The farm-fox experiment, 1–4.
Vohs, K. D., Finkenauer, C., & Baumeister, R. F. (2011). The Sum of Friends' and Lovers' Self-Control Scores Predicts Relationship Quality. *Social Psychological and Personality Science*, *2*(2), 138–145. http://doi.org/10.1177/1948550610385710

Walker, M. D., & Mason, G. (2012). Female C57BL/6 mice show consistent individual differences in spontaneous interaction with environmental enrichment that are predicted by neophobia. *Behavioural Brain Research*, *227*(2), 508–513. http://doi.org/10.1016/j.bbr.2011.09.032

Weinstein, T. A. R., & Capitanio, J. P. (2008). Individual differences in infant temperament predict social relationships of yearling rhesus monkeys, Macaca mulatta. *Animal Behaviour*, *76*(2), 455–465. http://doi.org/10.1016/j.anbehav.2008.01.024

Weir, A. A. S., Chappell, J., & Kacelnik, A. (n.d.). Shaping of Hooks in New Caledonian Crows, 11.

Wilsson, E., & Sundgren, P. E. (1998). Behaviour test for eight-week old puppies - Heritabilities of tested behaviour traits and its correspondence to later behaviour. *Applied Animal Behaviour Science*, *58*(1–2), 151–162. http://doi.org/10.1016/S0168-1591(97)00093-2

Chapter 2. DEVELOPMENT AND VALIDATION OF THE CANINE BODY LANGUAGE QUESTIONNAIRE (CBLQ)

Carolyn J. Loyer[1] & Robin L. Foster[1,2]
[1]University of Washington, Seattle, Washington, USA
[2]University of Puget Sound, Tacoma, Washington, USA

Abstract

Canine behavior is often quantified using owner reports of dog behavior. Because canine communication is largely visual and owners use visual cues to assess their dog's behavior, it is possible that important information is lost in the translation from visual processing to verbal reporting. To combat this potential problem, in Study 1 we developed the Canine Body Language Questionnaire (CBLQ), a pictorial questionnaire to assess dogs' body language on three dimensions: Bold, Aroused, and Aggressive. Sketches of dogs of a variety of pure- and mixed breeds in various positions were created and rated on each dimension by experts in canine behavior. We retained sketches with the highest inter-rater reliability, paired them by body type, and compiled pairings to represent each of the three dimensions (Bold, Aroused, and Aggressive). In Study 2, two hundred fifty-six dog owners completed the CBLQ as well as the previously validated Canine Behavioral Assessment and Research Questionnaire (C-BARQ), and we compared owners responses on both questionnaires. C-BARQ scores of Aggression and Fear were more tightly positively correlated with the proportion of Aggressive and High Arousal pictures owners selected on the CBLQ when only questions relating to dogs' behavior toward other dogs were used. The proportion of Bold pictures owners selected was negatively correlated with Aggressive, Aroused, and C-BARQ Dog-Directed Fear and Dog-Directed Aggression. In Study 3, a subset of the owners who completed the two questionnaires brought their dogs to participate in an experimental introduction to a live dog. Two raters scored dogs' behavior on

each dimension during the introduction to test for inter-rater reliability, which was high. We compared dogs' scores on these dimensions in the experimental context to the owners' answers on the CBLQ and found that the proportion of Aggressive pictures owners selected was positively correlated with observer ratings of Aggressive and a similar but non-significant trend for Aroused. There was no relationship between observer ratings of Bold and the proportion of Bold images owners selected. Our results demonstrate that these three proposed dimensions are not independent of each other, but do generally relate to owners' perceptions of their dogs' behavior in their daily lives. Further analyses should examine which features owners are using to identify dogs' body language, and whether those cues are reliable predictors of dog-dog aggression in the real world.

2.1 INTRODUCTION

Those who own or work closely with dogs can attest to the wide range of individual variation in their behavior in traits such as fear, anxiety, boldness, and aggression. An important advancement in animal personality research has been the development of standardized assessment scales to quantify these individual differences. Behavior assessments generally fall into two main types: direct observation of behavior in structured situations (behavioral assays) and questionnaires completed by owners about their dogs' behavior. Two canine behavior assays that are frequently used are the Safety Assessment for Evaluating Rehoming (SAFER®) developed by the American Society for the Prevention of Cruelty to Animals (ASPCA), and the Dog Mentality Assessment (Wilsson & Sundgren, 1997). The Canine Behavioral Assessment and Research Questionnaire (C-BARQ; Hsu & Serpell, 2003), on the other hand, is perhaps the most commonly used owner-completed canine personality questionnaire. The primary use of these assessments varies, with SAFER® often implemented as a screening tool for dogs that have been surrendered by their owners to shelters or rescue organizations, while the Dog Mentality Assessment is used in Sweden to identify ideal candidates to breed for particular tasks. The C-BARQ is frequently used as part of research investigating individual differences in canine behavior, from genetic foundations of aggression (Duffy, Hsu, & Serpell, 2008) to the relationship between owner and canine personality (Dodman, Brown, & Serpell, 2018). The large pool of C-BARQ surveys completed over the course of more than a decade enables scientists to detect even those variables that have a small effect size, as is common in genetic research (Plomin, 2013). The Dog Mentality Assessment is a battery of items developed and used in Sweden to inform breeding choices. See Wiener and Haskell (2016) for a comprehensive overview of canine behavior assessment scales.

There are benefits and detriments to each type of behavior assessment. Behavioral assays are time intensive, and require careful environmental control to maintain consistency across tests. Control and diligent training of evaluators performing behavioral assays may improve their objectivity compared to owner surveys, however the effort to standardize these tests may result in an environment that does not replicate real-world situations, making it challenging to interpret the results or apply them to predictions of dogs' behavior in their daily lives. By contrast, owner surveys allow a broad interpretation of a dog's behavior in a greater variety of situations than can practically be measured using an in-person assay, but may be more subject to individual differences in owners' interpretations of both dog behavior and the questions themselves. One major advantage of owner surveys is their ability to be administered to a broad sample in a short period of time, with minimal cost. This is especially important when evaluating factors that may relate to individual differences in canine behavior, but likely have a small effect size.

One facet of canine behavior that is commonly assessed in both behavioral batteries and owner surveys is sociability with conspecifics. Canine social interactions are mediated by olfactory, acoustic, and visual communication signals (Siniscalchi, d'Ingeo, Minunno, & Quaranta, 2018). Dogs are both senders and receivers in intraspecific signal exchanges, and owners may be viewed as eavesdroppers to these interactions. Humans are a visually-oriented species, thus dog owners tend to observe and attend to canine body language. The accuracy of owner reports of their dog's behavior or intentions towards other dogs, however, can be inconsistent or poor, and this may be in part due to the owners' limited ability to read, interpret or describe canine body language (Demirbas et al., 2016).

In the first case, the ability to read canine body language is limited either because the owner doesn't notice it or cannot see relevant cues. Owners may not notice subtle changes such as shifts in ear position, tail height or movement, weight distribution, or smaller mouth movements that precede more overt and escalated social cues (Wan, Bolger, & Champagne, 2012). Alternatively, owners may notice these small changes in body movements and postures, but fail to interpret or respond to them properly. Should this occur in a social interaction between dogs, it may lead to increased conflict because owners are unsure of whether or when to intervene. Finally, some owners may be able to read their dogs body language fairly well, but when asked to describe the behavior, they struggle to verbalize information that has been visually processed and encoded.

Owners may more readily and accurately describe their dog's body language when offered the chance to identify it in pictures instead of verbally. Verbal overshadowing has been shown to result in subsequent recognition errors of visual memory (Brown, Brandimonte, Wickham, Bosco, & Schooler, 2014), suggesting that asking owners to verbalize a visual memory may actually alter or shape that memory itself. Verbal descriptions may be particularly problematic when the individual has inadequate verbal knowledge to describe the memory of a nonverbal image. This general principle has been called the “modality mismatch assumption”, and is based on the principle that performance is best when memory encoding and retrieval overlap maximally (Brown et al., 2014). Given the visual nature of human observations of dog-dog interactions, as well as the practical and scientific significance of quantifying these interactions, we sought to create a questionnaire to allow owners to report their dog’s behavior with conspecifics, the Canine Body Language Questionnaire (CBLQ).

The first step in creating the CBLQ was to generate potential dimensions of body language that may be both important for predicting the outcome of an interaction, as well as visually identifiable in from a still image. Three dimensions of dog body language were examined, selected for their practical application and/or relevance to the scientific literature: Shy-Bold, Friendly-Aggressive, and Low-High Arousal. These represent three common ways of characterizing canine behavior, and indeed, behavior in many species. We did not assume that the three trait dimensions would be orthogonal and independent, and evaluating construct independence was a research goal of this preliminary scale development study.

The first dimension included in the CBLQ, Shy-Bold behavior, is repeatedly found in some form in the literature on individual differences in nonhuman animals, ranging from dumpling squid (*Euprymna tasmanica*) (Sinn, Apiolaza, & Moltschaniwskyj, 2006; Sinn, Moltschaniwskyj, Wapstra, & Dall, 2010) to great tits (*Parus major*) (Carere et al., 2004; Fidler et al., 2007) to capuchin monkeys (*Sapajus apella*) (Morton, Lee, & Buchanan-Smith, 2013). In dogs, boldness has been found to vary between breeds (Starling, Branson, Thomson, & McGreevy, 2013), to be heritable (Strandberg, Jacobsson, & Saetre, 2005), and to influence lower-order dimensions of canine personality (Svartberg, 2005; Svartberg & Forkman, 2002). Thus, we wanted to investigate whether this potentially important dimension could be captured in a pictorial questionnaire.

Behavior on the shy-bold personality dimension likely influences canine social interactions, and we predicted that owners would be aware and easily recognize if their dog was fearful (shy) or confident (bold). Fear is a universal emotion in animals (Panksepp, 1998, p. 206), and past research has linked fearful dogs with a risk of aggression. For example, when classifying the types of aggression seen in dogs that presented at a behavior clinic, Borchelt (1983) found that aggression elicited by fear was the most prevalent, thus we might expect that agonistic interactions are more likely when at least one partner is fearful. If true, compared to confident (bold) dogs, fearful (shy) dogs should have a greater incidence of aggression and, more generally, may engage in poorer quality social interactions. Therefore, we expected that fear and degree of sociability may be related.

The second dimension of the CBLQ, Friendly-Aggressive, directly targets whether a dog's interactional style with other dogs is prosocial or socially antagonistic. Being able to determine whether a dog is soliciting a positive interaction or is threatening a negative interaction is both biologically relevant and of practical importance to dog owners looking to prevent aggressive interactions. Evidence suggests that although humans are generally good at categorizing dogs that are "happy", they may struggle to accurately identify signs of fear without advanced experience or training in canine behavior (Tami & Gallagher, 2009; Wan et al., 2012). This difficulty may be further exacerbated when there is ambivalence in a dog's motivation to engage in a social interaction.

Approach-avoidance conflict is often seen in social contexts because the opportunity to socialize holds the potential to be rewarding or punitive (McNaughton & Corr, 2008, p. 72), in dogs as with any social species. Social interactions may be positively valenced, as when strengthening or consolidating bonds or alliances, or attracting a mate. They may also be negatively valenced, as in the case of competition over resources or in aggressive interactions that occur when rival groups meet or when a dominance hierarchy is in flux. Emotional/motivational valence is a complicated construct, however, the Friendly-Aggressive dimension aims to explore the valence of social interactions. Of the three potential dimensions examined in this study, it is the most applied, seeking to answer the questions; “Is your dog friendly?” and “Is that a dangerous aggressive dog?”. Each of these questions are of great concern to dog owners, applied dog behavior specialists, the general public, and the legal system.

The final dimension, Low - High Arousal, addresses an individual’s tendency to be alert and prepared to take action, and may mediate how readily that individual will approach (or avoid) a conspecific in a social situation (Corr, 2008, p. 14), as well as the degree to which an individual is able to incorporate peripheral cues into their decision-making (Easterbrook, 1959). Physiologically, preparedness for a real or perceived crisis involves activation of the sympathetic division of the autonomic nervous system, the classic 'fight or flight response, with increased blood flow, skeletal muscle tension, and respiration. At the same time, circuits in the Central Nervous System inhibit the parasympathetic nervous system (Bear, Connors, & Paradiso, 2001, p. 507). Mentally, arousal is generally associated with activation of the reticular activating system and cortical brain areas (Corr, 2004; Eysenck, 1976). Whereas individuals vary in their degree of arousal across contexts, there is also evidence that individual differences in arousal underlie common personality factors such as Extraversion-Introversion (Stelmack, 1990).

In human emotion and motivation research, arousal (low - high) and valence (pleasant - unpleasant) are the only two dimensions in the Emotion Circumplex model of Affect (Russell, 1980). This model visually displays the relationship of human emotional states on a two-dimensional axis, with arousal as the Y-axis mediating emotional activation. In dogs, physical activity and reactivity (or over-reactivity) to stimuli are the most commonly studied correlates of arousal, and are frequently identified as important ways in which individuals vary (Beaudet et al., 1994; Bradshaw & Goodwin, 1998; Campbell, 1972; Hsu & Serpell, 2003; Goddard & Beilharz, 1984b; Coren, 1995; Serpell, 1983; Weiss & Greenberg, 1997). Researchers have repeatedly demonstrated that arousal can influence behavior and decision-making (Bray, MacLean, & Hare, 2015; Easterbrook, 1959; Rogers, Lancaster, Wakeley, & Bhagwagar, 2004). The Reinforcement Sensitivity Theory (RST; Corr, 2008; McNaughton & Gray, 2000) is a comprehensive theory that proposes environmental and neurological pathways for arousal and behavior to interact.

The RST identifies three distinct brain-behavioral systems, including the 'Fight-Flight-Freeze System' (FFFS, activated in response to threatening stimuli), the 'Behavioural Approach System' (BAS, activated by rewarding stimuli) and the 'Behavioural Inhibition System' (BIS, triggered when the FFFS and BAS are both highly active). These distinct systems reflect individual differences in sensitivity to punishment, reward, and mediation between the two. In research with humans, RST has been used to predict anxiety, impulsivity, and extraversion (Corr & McNaughton, 2008). Dogs similarly display individual differences in anxiety, impulsivity, and extraversion, therefore RST may provide a useful framework for canine personality research.

According to RST, arousal increases and decision-making decreases when FFFS and BAS are simultaneously activated, leading to conflicting avoid and approach motivations; when the motivations are relatively equal, the BIS is activated, leading to increased arousal and risk aversion (Corr & McNaughton, 2008, p.166). Arousal may mediate behavior based on the other two states, for example by influencing how quickly dogs switch between FFFS and BAS. Therefore, we expected that arousal may not be independent of the other two dimensions, and that a dog's degree of fear may also correlated to the level of sociability (friendliness versus aggression) expressed in their body language.

In three separate studies, we first developed a pictorial questionnaire to evaluate dogs' body language when interacting with other dogs, then compared responses on the questionnaire to owner responses on the C-BARQ, and finally to dogs' behavior in an experimental introduction to another dog.

2.2 Study 1: Scale Development

2.2.1 Methods

Item generation: Creating sketches

We collected still images of dogs from professional and amateur dog photographers and from an internet photograph search. In order to be included, an image had to depict a dog's full body from a lateral or ¾ view, without obstruction of the dog's body by other dogs, people, or objects. Images were also taken from frames from videos of dogs interacting, if they met the criteria above. We gave preference to dogs interacting with or in the presence of other dogs as these images expressed canine body language in a real-life social context. Photographs were selected to ensure representation of the full range on each of the three personality dimensions investigated.

Sketches were created by tracing the outlines of silhouettes with simple internal structures from the 2-D photograph images using pencil and paper on a light board, and the sketches were then scanned into .jpg files (Figure 1). We inspected the sketches for contrast and clarity, and corrected them when necessary to ensure visibility in an online questionnaire. In addition, for each sketch we catalogued specific body language features considered important in canine social communication, such as tail position, ear position, posture, and mouth shape, (Wan et al., 2012). We also recorded physical characteristics that we could potentially impact the saliency of these features, including tail length, coat length and texture, ear type, breed type, etc. In total, seventy-eight sketches of various breeds and sizes of dogs displaying a range of body postures were compiled to be sent to a panel of experts.

Sketches have some advantages over rich, textured, and detailed photographic images as visual stimuli in research. The simple line drawings can hone in on canine body language features that are typically attended to and learned by owners, and thereby reduce challenges of 2D photographs such as viewpoint variation, partial occlusion of the image, low-resolution and motion-blur (Shi, Hospedales, & Xiang, 2015). In one study using MRI to compare visual recognition of photographs and line sketches of natural scenes (Walther, Chai, Caddigan, Beck, & Fei-Fei, 2011), researchers found that despite the loss of detail, line sketches that retained global structure features were classified as readily as were color photographs of the scenes. The authors concluded that detail (local structure) was not necessary for accurate categorization when global structure of the scene was retained in the line sketches(Walther et al., 2011).

Item generation: Expert screening

We recruited professional dog behavior consultants via word of mouth and via email lists for professional organizations that offer certifications in canine behavior or training. Diesel, Brodbelt, and Pfeiffer (2008) found that increased levels of training and experience improved the reliability of behavioral scoring by shelter staff. Involving experts in the scale development item generation process is standard practice, and in our study we assumed that, compared to novice dog owners, professionals working with dogs would have the knowledge and experience to recognize and judge both global and local features of the sketches that might indicate the behavioral dimensions we were interested in assessing. This was a particularly important step in order to ascertain the validity of using sketches to capture canine body language. As noted by Walther et al. (2011) "the line drawing of an image is a drastically impoverished version of the picture (p. 9665), and in the process of simplification our sketches may have lost essential detail for image recognition and discrimination.

Six experts with five or more years of professional experience working with dogs (M = 23.43, SD = 11.50) completed ratings for all pictures. Each of the six dog behavior experts was certified through a professional animal behavior or dog training organization; three were Certified Applied Animal Behaviorists (CAAB) through the Animal Behavior Society, one was an associate CAAB (CAAB certification for those with master's level education), four had previously been or were currently Certified Professional Dog Trainers – Knowledge Assessed through the Association of Professional Dog Trainers, and one was a leading expert in dog body language with experience in dog training, behavior consultation, behavior evaluations, and having given professional talks and developed a mobile app to educate owners about dog body language.

Preliminary questionnaire development

The preliminary questionnaire was created and distributed using a web-based platform (Qualtrics, Provo, UT), and was divided into four sections. The first section asked questions regarding respondents' experience and training in canine behavior to verify their eligibility to participate. In the subsequent three sections, respondents viewed the sketches of dogs and were asked to rate the dogs' body language on three different scales, one scale per section: Bold (Shy-Bold), Aroused (Low-High), and Aggressive (Friendly-Aggressive). Respondents viewed one image at a time in a randomized order, and rated body language on a 5-point Likert scale. This was repeated for each scale, so that respondents rated each image three times. On the preliminary questionnaire, the 5-point Likert scale anchor points were as follows: Bold was rated 1=Shy, 5=Bold, and 3=Neutral/ambiguous. Sociability was rated 1=Friendly, 5=Aggressive, and 3=Neutral/ambiguous. Arousal was rated 1=Low arousal/relaxed, 5=High arousal/excited, and 3=Moderate arousal. No specific instructions were provided for rating the Bold scale. In the instructions for rating sociability, "friendly" was defined as "soliciting social interaction" and "aggressive" was defined as "discouraging social interaction." In the instructions for rating arousal, we specified that respondents were to focus on the physiological states of arousal such as increased muscle tension and/or heart rate.

2.2.2 Results and Discussion

Image selection for scale items

We calculated the correlation between the experts' ratings of each sketch on the three scales. The ratings for Bold and Aroused were weakly negatively correlated (r = -0.332, p = 0.0030), such that sketches of dogs that experts rated as bolder tended to be rated as less aroused. Bold was moderately negatively correlated with Aggressive (r = -0.583, $p = 2.19 \times 10^{-8}$), such that sketches of dogs that experts rated as bolder were also rated as more friendly. Aggressive and Aroused were moderately positively correlated (r = 0.517, p = 1.286 x 10^{-6}), such that sketches of dogs that experts rated as more aggressive tended to be rated as more highly aroused.

The negative relationship between the Bold and Aggressive scales makes ecological sense, particularly given our descriptions for the friendly (soliciting social interaction) and aggressive (discouraging social interaction) anchors. Body language associated with fear often includes features indicating avoidance (e.g. turning away, refraining from eye contact; (Döring, Roscher, Scheipl, Küchenhoff, & Erhard, 2009). These social signals have been proposed by Turid Rugaas as a means of de-escalating conflict or agonistic interactions, and serve to discourage further (negative) social interaction. Therefore, it is in an individual's interest that fear-based signals be unambiguous in order to maximize their effect on the social partner. Conversely, when an individual is more confident in a positive outcome of the social encounter, whether in anticipation of a pro-social interaction or an agonistic one in which they are likely to prevail, there may be fewer advantages and perhaps even some costs to broadcasting exaggerated social communication signals.

A primary goal of the preliminary questionnaire was to identify which sketches to retain for use in the final questionnaire. To evaluate interrater agreement of the experts, we calculated the standard deviation (SD) and interquartile range of their ratings for each sketch on the three personality dimension scales. The smaller the SD of the average ratings across the six experts, the more similar the ratings of the sketches were across experts. Sketches were retained only if the SD of the experts' ratings was at or below the first quartile (see SD cut-off, Table 1). Based on this criterion, 19 (24.36%) of the 78 sketches had sufficient agreement to be included on the

Bold scale, 32 (41.03 %) had sufficient agreement for the Aggressive scale, and 17 (21.79%) had sufficient agreement for the Aroused scale.

From the pool of high interrater agreement items, we paired sketches of dogs demonstrating distinct ratings on each scale (e.g. “bold - neutral”, “low arousal - high arousal”, “neutral – friendly”). Each scale also included one or two pairings of “neutral - neutral” sketches so that in future analyses we could potentially identify which features of body language were most salient, and which features owners attended to most or recalled best. For example, if sketches of two dogs rated as “neutral/ambiguous” on arousal were paired, did owners tend to select the dog based on the position of their ears, tails, or mouths, or their weight distribution? Because we anticipated that breed differences in physical appearance might take precedence over body language in terms of owners’ tendency to identify a sketch with their own dogs, we matched sketches as closely as possible by appearance and body type. For instance, sketches of dogs with cropped tails, short coats, and long snouts were paired, while sketches of dogs with short tails and curly coats were paired. When possible, dogs were matched by exact breed. Feedback from the expert raters indicated that it was difficult to interpret body language in the absence of the social context. Therefore, in the questionnaire for owners, the sketches were positioned so that the two dogs depicted were facing each other as if interacting.

The revised CBLQ introduced in studies 2 and 3 (below) included twenty-eight pairs of sketches across the three scales (see Table 2 for exact counts of each type of pairing). For each item pair a forced-choice format was applied, and owners completing the questionnaire were asked to select one of the two sketches that best matched their own dog when interacting with other dogs. We avoided using a sketch more than once in the questionnaire. Fifty one sketches were unique, but five sketches were included twice in different image pairs. Two sketches were used twice on the same scale dimension. One sketch was used twice in the Bold dimension, one in the Aggressive dimension. All sketches used for the Aroused dimension appeared only once in that dimension.

Scoring for the revised CBLQ was designed such that higher total scores indicated that a dog's body language was Bold, Aggressive, or High Arousal on the respective scales, and lower total scores indicated Shy, Friendly, and Low Arousal body language. To calculate an overall score on the Bold scale, for each pair of sketches, the one depicting body language higher on the bold continuum was assigned a "1" and the sketch lower on the bold continuum was assigned "0". The maximum cumulative raw score for the Bold scale was 6, indicating the sketch depicting the bolder dog had been selected for all of the items. The same system was applied to the Aggression and Arousal scales, with a total raw score of 9 on each scale. Selections for the "neutral-neutral" pairings were not included in the tally. To standardize the scores across dimensions, raw scores were converted to proportions. A copy of the C-BLQ can be found in Appendix 1.

2.3 Study 2: Item analysis and convergent validity with C-BARQ

2.3.1 Methods

Subjects

To determine whether the pairs of sketches resulting from Study 1 related to other measures of canine behavior, we recruited dog owners via word of mouth and social media to complete a battery of questionnaires. This battery included questions about the dogs' demographics and history, along with the CBLQ and the C-BARQ (see Appendix 2 for the complete survey).

Two hundred fifty-six dog owners completed the survey. The average age of the dogs was 5.22 $\pm$ 3.05 years, and they ranged in age from 6 months to 13.7 years. The average duration the dogs had been living with their current owners was 3.92 $\pm$ 2.92years, with a range of 1 month to 13.3 years. Of the dogs included, 126 were females (49.22%) and 128 were male. Of the females, 4 (3.17%) were intact, 110 (87.30%) were spayed, and 12 (9.52%) had unspecified neuter status. Similar proportions were found in the males, with8 (6.25%) intact, 106 (82.81%) neutered, and 14 (10.94%) with unspecified neuter status. Sixty-eight owners reported that their dogs were purebred, either purchased from a breeder or with AKC registration, and 23 believed their dogs were purebred based on appearance. Fifty-one owners reported with confidence that their dog was a mixed breed, 105 owners guessed their dog's breed based on physical appearance, and 5 owners did not know their dog's breed mixture.

Statistics

We used the psycho package in RStudio to test for pairwise correlations between the owner answers on the CBLQ and their answers on the C-BARQ (see Table 3 for a summary of each C-BARQ scale), using the Holm-Bonferroni method to control for multiple comparisons. Given that Study 1 showed correlations in expert ratings between the three scales, we expected some correlations between the three dimensions for owner responses as well, assuming that owners attend to, recall, and identify similar behavior patterns as the experts.

Comparing owner responses on the CBLQ and on the C-BARQ allowed us to examine both the discriminant and the convergent validity of the CBLQ. If, as intended, owner responses on the CBLQ reveal dogs' behavior during dog-dog interactions, we should see weak (or no) correlations between CBLQ scores and dogs' scores on unrelated C-BARQ scales, such as dogs' tendency to display signs of anxiety when separated from their owner (separation-related behavior, or SRB), or Trainability, which measures dogs' responsiveness to their owner's obedience cues. Conversely, C-BARQ scales such as Aggression (the average severity of a dog's aggressive response ranging from 1 = No aggression to 5 = Severe aggression), Fear and Anxiety (Fear; the average severity of a dog's fearful responses ranging from 1 = No fear to 5 = Extreme fear), and Excitability (a dog's tendency to become highly active in a variety of situations such as when the owner returns home after a short absence, or prior to going out for a walk) should be related to the CBLQ dimensions. Further, the correlations between the CBLQ scores and C-BARQ Aggression and Fear scales should be stronger when including only those questions from the scales that relate specifically to interactions with other dogs.

2.3.2 Results and Discussion

Correlations between CBLQ dimensions

The proportion of Aggressive sketches and the proportion of Bold sketches that owners selected were moderately negatively correlated (r = -0.43, p = 5.663 x 10^{-13}). Owners that tended to select the bolder of two sketches also tended to select the more friendly of two sketches. The proportion of Aggressive and High Arousal pictures owners selected were moderately positively correlated (r = 0.47, p = 2.517 x 10^{-15}). Owners who chose a high proportion of sketches that experts rated as more aggressive also chose more High Arousal sketches. The relationship between the proportion of Bold sketches selected and High Arousal sketches selected trended toward a weak, negative correlation (r = -0.186, p = 0.00285), but was not significant after controlling for multiple comparisons (see Table 4 for the full table of questionnaire scoring correlations).

Correlations between the CBLQ and C-BARQ

Bold was moderately negatively correlated with C-BARQ Fear ($r = -0.35$, $p < 0.001$) and DDFear ($r = -0.35$, $p < 0.0001$), and weakly negatively correlated with C-BARQ Aggression ($r = -0.19$, $p < 0.05$), Excitability ($r = -0.21$, $p < 0.05$) and SRB scores ($r = -0.21$, $p < 0.05$). Owners that chose a higher proportion of Bold pictures in the CBLQ tended to report that there dog was less fearful, less aggressive, less excitable, and shows fewer signs of anxiety when left alone.

Aggressive was moderately positively correlated with C-BARQ DDAggr ($r = 0.40$, $p < 0.001$), DDFear ($r = 0.36$, $p < 0.001$), and C-BARQ Aggression ($r = 0.35$, $p < 0.001$). Aggressive had a weak, negative correlation with C-BARQ Fear ($r = 0.25$, $p < 0.01$). Owners that reported that their dog is more aggressive, especially toward other dogs, and more fearful toward other dogs tended to select more Aggressive pictures on the CBLQ. Owners reported higher generalized Fear on the C-BARQ, however, tended to select fewer Aggressive pictures on the CBLQ.

Aroused (High arousal) was moderately positively correlated with C-BARQ DDAggr ($r = 0.48$, $p < 0.001$), Aggression ($r = 0.47$, $p < 0.001$), and DDFear ($r = 0.37$, $p < 0.001$). There was a weak positive correlation between Aroused and C-BARQ Fear ($r = 0.22$, $p < 0.05$), Excitability ($r = 0.19$, $p < 0.05$), Attachment ($r = 0.20$, $p < 0.05$), and Miscellaneous ($r = 0.19$, $p < 0.05$). Dog owners that selected more High arousal pictures on the CBLQ tended to report on the CBARQ that their dogs were more aggressive overall as well as with other dogs, show more severe fear with other dogs and in general, and were more excitable, more attached to their owners, and show more miscellaneous behavior challenges such as fixating on light or chasing squirrels.

The results of our pairwise comparisons suggest that the dimensions measured in the CBLQ are most strongly correlated with each other, and with the C-BARQ scales of Fear and Aggression. The dimension of Aggression on the CBLQ showed the highest specificity, relating only to the overall scales of Aggression and Fear on the C-BARQ, and their subsets including dog-directed questions. This provides strong support for the conceptual validity of using a pictorial questionnaire to assess the valence of dog-dog social interactions (prosocial or antagonistic). The CBLQ dimension of Arousal showed the least specificity in terms of focusing on dog-dog interactions. If dogs that are highly aroused in dog-dog interactions tend to be generally excitable, they may also show more extreme (and problematic) responses to a variety of situations, not solely in their reactions to conspecifics.

Principle Components Analysis of the CBLQ and C-BARQ scores

To obtain a more global understanding of how owners' answers on the two questionnaires were related, we conducted principle components analysis (PCA) using scaled CBLQ and C-BARQ scores using the factoextra package in R. A primary objective of this study was to evaluate the validity of the CBLQ as an assessment of social interactions between dogs. To this end, we conducted two separate analyses: The first analysis used the complete set of questions from all scales on the C-BARQ as well as the CBLQ. The second analysis used subsets of questions on the C-BARQ Fear & Anxiety and Aggression scales that pertained to dogs' reactions toward other dogs, labelled Dog-Directed Fear (DDF) and Dog-Directed Aggression (DDA). In this second analysis the DDF and DDA scores replaced the complete Fear and Aggression scores used in the first analysis. These scale subsets have been used in previous work to distinguish between the targets of fearful and aggressive behavior for the purpose of determining cognitive (Barnard, Wells, Hepper, & Milligan, 2017), genetic (Liinamo et al., 2007), and environmental factors (Duffy et al., 2008; Hsu & Sun, 2010) associated with these behavior problems.

Comparing the results of these two analyses gave us an opportunity to examine whether owner responses on the CBLQ adequately targeted dogs' body language during dog-dog social interactions, as intended, or if owners selected sketches based on a broader memory and mental representation of their dog's body language in all situations. If the CBLQ is a good tool for assessing dog personality, we would expect to see related loadings between the CBLQ dimensions and C-BARQ Fear and Aggression scales, but would not necessarily expect the CBLQ dimensions to load with other C-BARQ scales such as SRB or Attachment. In addition, if the CBLQ discriminates body language displayed during intraspecifc encounters from body language displayed in other contexts, we should expect more tightly correlated results when using questions that only relate to dog-dog interactions on the C-BARQ Fear and Aggression subscales than when using the complete scales.

PCA using full C-BARQ scores

Visual examination of the scree plot (Figure 2a) indicated that the variance was adequately captured in two components. The first component contained positive loadings for C-BARQ Aggression (0.399), Fear (0.359), Excitability (0.348), SRB (0.313), Attachment (0.304), and Miscellaneous (0.363). The proportion of Aggressive pictures owners selected on the CBLQ and the proportion of High Arousal pictures they selected also loaded positively (0.295 and 0.299, respectively), whereas the proportion of Bold pictures selected loaded negatively (-0.280). The second component contained stronger positive loadings for Aggressive (0.556) and Aroused (0.375), and a negative loading for Bold (-0.296), while C-BARQ Aggression and Fear scores did not load significantly on the second dimension. Excitability, SRB, Attachment, Trainability, and Miscellaneous all loaded negatively on the second dimension (Figure 2b).

PCA using Dog-Directed Aggression & Fear scores

There was one primary change in results when the PCA was re-run using only the dog-directed subsets for the C-BARQ Aggression and Fear scales. Similar to the PCA with the full C-BARQ Aggression and Fear scale, the second analysis was also adequately captured in two components (Figure 3a). However, the loadings for DDAggr and DDFear were more tightly correlated with the CBLQ scores PropAggr and Prop High, and still in opposition to PropBold (Figure 3b). All other C-BARQ scores maintained equivalent loadings to those in the first analysis. In summary, the scores on the CBLQ dimensions were most strongly and consistently related to other CBLQ scores, and to the C-BARQ scales of Fear and Aggression. When using the subsets of questions on these scales that focused on dog-dog interactions, CBLQ scores showed stronger correlations than with the complete scale, suggesting that the CBLQ is targeting canine behavior in dog-dog interactions.

These findings suggest that not only are the owner answers on the CBLQ related to the scales on the previously validated C-BARQ, but that owners are responding in a way that emphasizes their dog's behavior toward other dogs, and not based on a memory of their dog's global behavior. The loadings for C-BARQ's Attachment and Attention-Seeking, SRB, and Miscellaneous scores loaded onto the dimensions identified by PCA distinctly from the CBLQ dimensions and from C-BARQ Fear and Aggression scales, providing evidence of the discriminant validity of the CBLQ. Combined, our results indicate that the use of a pictorial questionnaire for assessing body language may be a useful addition to current behavior questionnaires.

2.4 Study 3-Construct validity with participant dogs greeting a live dog

2.4.1 Methods

The participants in Study 3 were part of the larger group of dog owners who were recruited for Study 2. They included 25 owners of dogs of various breeds who brought their dogs to participate in-person during an experimental structured set up where the dogs were introduced to and approached four stimuli (a live dog, a fake stuffed model dog, a novel object, and a control condition). The stimuli were presented one at a time in a counterbalanced order and the trials were video recorded for later behavioral coding. All four stimuli were introduced during a single session. See Loyer et al. (2019, in prep) for details of the participation.

Two independent observers scored the global behavior of participant dogs from video recordings on the three dimensions, using the same 5-point Likert scale used by experts to rate the original sketches (see Study 1). Inter-rater reliability was calculated on scores for live dogs on the three dimensions, Bold, Aggressive, and Aroused. One rater (CL) also scored a subset of dogs twice to assess intra-rater reliability.

Observers also coded the participant dogs' body language more specifically. They rated average body position (weight forward, neutral, or back), tail position (the height of the base of the tail in relation to the spine, ranging from 1 = tail tucked, to 5 = base of tail at or near 90 degrees from the base of the spine), mouth position (open, closed, or in motion), and ear position (ranging from 1 = low and back to 5 = high and forward) at three different distances from the stimuli (32', 20', and 8'), as well as when given an opportunity to interact with the stimuli. One observer (CL) used the same body language coding to rate the body language of dogs depicted in the CBLQ sketches. For future work, we will compare average ratings of live dog body language scores as described above to the comparable scoring of the CBLQ line sketches owners identified as matching their dog's body language.

We calculated the Pearson's r correlation between observer ratings of dogs' behavior on the three scales during their introduction to the live dog with the owner reports via the CBLQ. The raters were unfamiliar with the dogs, but were specifically rating the dogs on the three scales of interest. The opposite was true of owner scores on the CBLQ: owners were very familiar with their own dogs, but blind to the three dimensions we were measuring.

2.4.2 Results and Discussion

The scores assigned by two independent observers to participant dogs (N=24) on the personality dimensions Bold, Aggressive, and Arousal (Table 5) were significantly correlated in all stimulus conditions, confirming high inter-rater agreement when these traits were measured on a 5-point Likert scale. Compared to Observer 1, Observer 2 consistently (in 11 of 12 stimulus-dimension conditions) scored dogs as bolder, more aroused, and less aggressive.

To assess the relationship between owner reports of dogs' behavior with conspecifics in their daily lives and the dogs' behavior in the experimental context, we compared observer ratings of the dogs' behavior in the live dog condition to owner responses on the CBLQ (see Table 6 for full correlation matrix). There was a moderate, positive correlation between observer ratings of aggression and owner scores on Aggressive on the CBLQ. Dogs whose owners selected a greater proportion of aggressive pictures on the CBLQ to represent their behavior toward other dogs tended to be scored as more aggressive during their introduction to a live dog ($r = 0.57$, $p = 0.0038$). The relationship between observer ratings and CBLQ scores on Aroused trended positive ($r = 0.36$, $p = 0.086$). There was no relationship between the proportion of Bold pictures owners selected and observer ratings of dogs' boldness during their introduction to a live dog ($r = 0.13$, $p = 0.54$, Table 5).

Given the pervasive presence of the shy-bold dimension in research in individual differences across taxa, the fact that in this study we did not find a relationship between owner responses and observer scores is of particular interest. It is possible that this is due to the way in which dogs were introduced in the experimental set up. During each trial, there was a chain-link fence between the subject and the stimulus dog. Since many dog-dog greetings occur while dogs are on walks with their owners, most often a leash is the primary barrier to physical interaction between dogs (as opposed to a chain-link fence), so behavior observed in our experiment may have differed as a result. Further, the stimulus dog was selected for her ability to remain still and calm during trials, and was offered treats throughout to maintain attention away from the subject dogs. These efforts to improve experimental control may have reduced the level of threat presented for those subjects that tended to be fearful of other dogs. However, given that we still observed a relationship between observer and owner ratings on Aggressive, and a trend towards a relationship on the Aroused dimension, our results suggest that Bold was the least robust dimension using our metrics.

The parallel loadings of Bold and C-BARQ Trainability in the PCA analyses from Study 2 also present an unexplained relationship. Although not significantly correlated on their own, their loadings were in the opposite direction from every other scale on the first dimension. Being "bold" and being "trainable" are both generally positive personality traits, but their definitions in the literature are quite variable particularly when compared to other measured behavior facets such as exhibiting separation-related distress or showing aggression towards other dogs. It may be that in order to improve the CBLQ and its interpretation, considering alternative dimensions is advisable.

2.5 CONCLUSION

Our results from Studies 1 and 2 indicate that the dimensions of body language selected for the CBLQ are not independent: Body language associated with or interpreted as boldness is also associated with being friendly and with a state of lower arousal, particularly in the context of dog-dog interactions. Feedback from experts who rated the original sketches on the three dimensions also indicates that body language is difficult to interpret when presented in a static image, separate from the context that generated it. This may be especially true for social interactions, since body language from one partner may be a response to the other partner's cues. In Study 2, owners were asked to select which of a pair of sketches best matched their own dog's body language, thus providing some social context for the body language displayed. The results from this study, comparing owner responses on the CBLQ to the C-BARQ, demonstrated that owners were responding based on their memory of their dog's behavior during interactions with conspecifics, therefore discriminating this from their dog's more global body language.

Study 3 demonstrated substantial inter-rater reliability when rating videos of the dogs on the three proposed dimensions, but less agreement between ratings in an experimental context and owner selections based on the dogs' behavior in their daily lives. This may be due to the difference in context, or may suggest that more careful description of the meaning of these dimensions is required. The finding that observer ratings and owner selections for the Bold dimension were not significantly correlated was surprising and of some concern, given its pervasive presence in scientific literature on individual differences. This dimension, in particular, may require a more specific operational definition to improve its consistency and validity.

Overall, our findings support the use of a pictorial questionnaire to evaluate canine behavior. In future studies, we hope to investigate which aspects of dogs' body language owners and experts primarily attend to when matching their dog's behavior or rating dogs' behavior, respectively. Future iterations of the CBLQ should remove duplicate images, and perhaps incorporate the expert ratings of each image on each dimensions when calculating owner responses to gain a more nuanced measure of dogs' body language from the questionnaire.

2.6 Tables and Figures

Figure 1. Exemplars of images used in the CBLQ. Photographs used to create the CBLQ (on the left), matched with the resulting sketches (to the right). 78 such sketches were sent to experts to rate on the three dimensions of the CBLQ: Bold, Aroused, and Aggressive. Top left image, retrieved from sites/87/2016/09/Canine-Body-Language.pdf

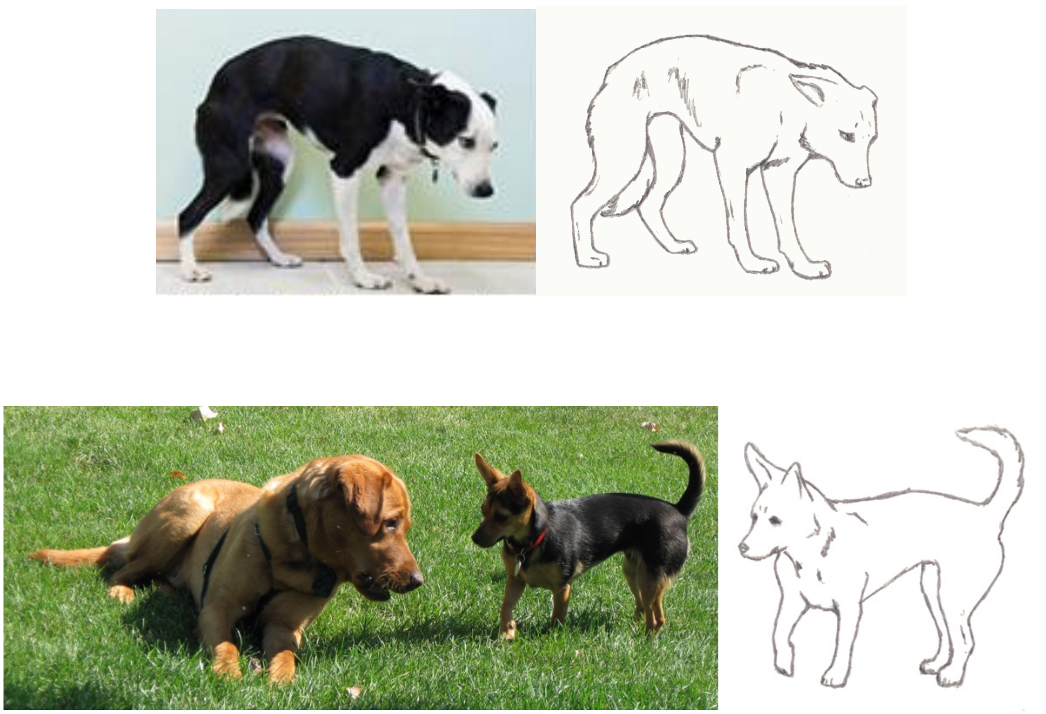


Table 1.
Standard deviation(SD) of expert ratings of sketches on each scale

Scale	**Median SD**	**Min SD**	**Max SD**	**SD cut-off**
Bold	0.58	0	1.12	0.471
Aroused	0.748	0	1.55	0.490
Aggressive	0.490	0	1.10	0.40

Note. The cut-off value indicates the SD at which ratings were considered too varied for a sketch to be used to represent that particular dimension.

Table 2.
The number of pairs used for each CBLQ scale

Counts of pairings					
Bold - Shy		**Aggressive - Friendly**		**High Arousal vs. Low Arousal**	
Bold vs. Shy	2	Aggr vs. Friendly	3	High vs. Low	3
Bold vs. Neutral	2	Aggr vs. Neutral	3	High vs. Neutral	3
Shy vs. Neutral	2	Friendly vs. Neutral	3	Low vs. Neutral	3
Neutral vs. Neutral	1	Neutral vs. Neutral	1	Neutral vs. Neutral	1

Table 3.
C-BARQ scales and descriptions

Scale	Description
Trainability	Responsiveness to owner cues, sensitivity to potential rewards and aversives. Higher score = More "trainable"
Aggression	Severity of aggression displayed in a variety of situations, e.g when verbally or physically corrected, during grooming, etc. Higher score = more severe aggression
Dog-Directed Aggression (DDA)	Severity of aggression displayed during interactions with other dogs. Higher score = more severe aggression
Fear and Anxiety	Severity of fear displayed in a variety of situations, e.g. when hearing a vacuum cleaner or fireworks, or when an unfamiliar human approaches. Higher score = more severe fear
Dog-Directed Fear (DDF)	Severity of fear displayed during interactions with other dogs. Higher score = more severe fear
Separation-related behavior (SRB)	Frequency of particular signs of distress are displayed when the dog is left alone. Higher score = more frequent SRB
Excitability	Degree of reaction to a variety of situations, e.g. when the doorbell rings or before a car ride. Higher score = more excitable
Attachment and Attention-seeking	Frequency with which dog follows or solicits attention from a particular human household member. Higher score = more attachment
Miscellaneous	Frequency of other potential behavior concerns, e.g. eating feces, excessive licking. Higher score = more frequent behaviors.

Table 4.

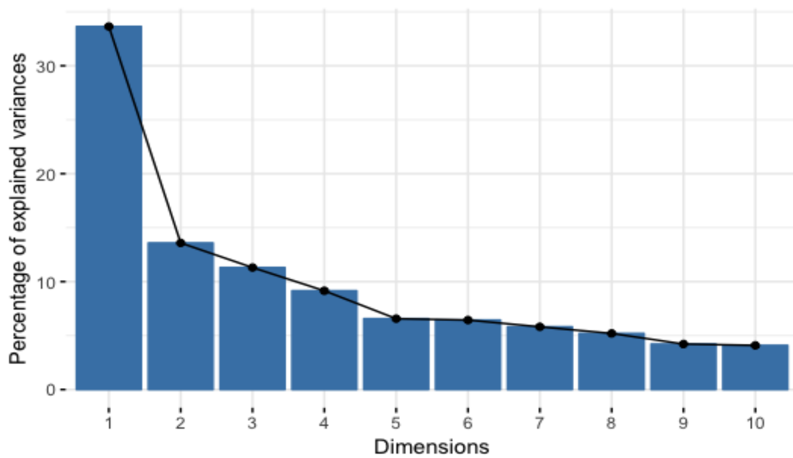
Correlation matrix of the relationships between CBLQ and C-BARQ dimensions

	Bold	Aggr	High	DDAggr	Aggr	DDFear	Fear	Excit	Train	SRB	Attach
Bold											
Aggr	-0.43***										
High	-0.19	0.47***									
DDAggr	-0.11	0.4***	0.48***								
Aggr	-0.19*	0.35***	0.47***	0.69***							
DDFear	-0.35***	0.36***	0.37***	0.39***	0.31***						
Fear	-0.35***	0.25**	0.22*	0.16	0.38***	0.61***					
Excit	-0.21*	0.14	0.19*	0.21*	0.43***	0.23**	0.32***				
Train	0.1	-0.17	-0.06	-0.15	-0.12	-0.09	-0.19	0			
SRB	-0.21*	0.13	0.12	0.11	0.25**	0.19	0.37***	0.32***	-0.16		
Attach	-0.19	0.14	0.2*	0.28***	0.35***	0.21*	0.23**	0.4***	0.08	0.3***	
Misc	-0.19	0.14	0.19*	0.26**	0.42***	0.24**	0.36***	0.43***	-0.12	0.43***	0.33***

Note: CBLQ dimensions are indicated in green, C-BARQ scales are indicated in purple. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 2. Results of Principle Components Analysis of CBLQ and C-BARQ full scale scores. a) Scree plot of the PCA indicating that 2 components are sufficient to explain the variance in scores, and b) Plot of the C-BARQ and BLQ loadings on the two components.

a)



b)

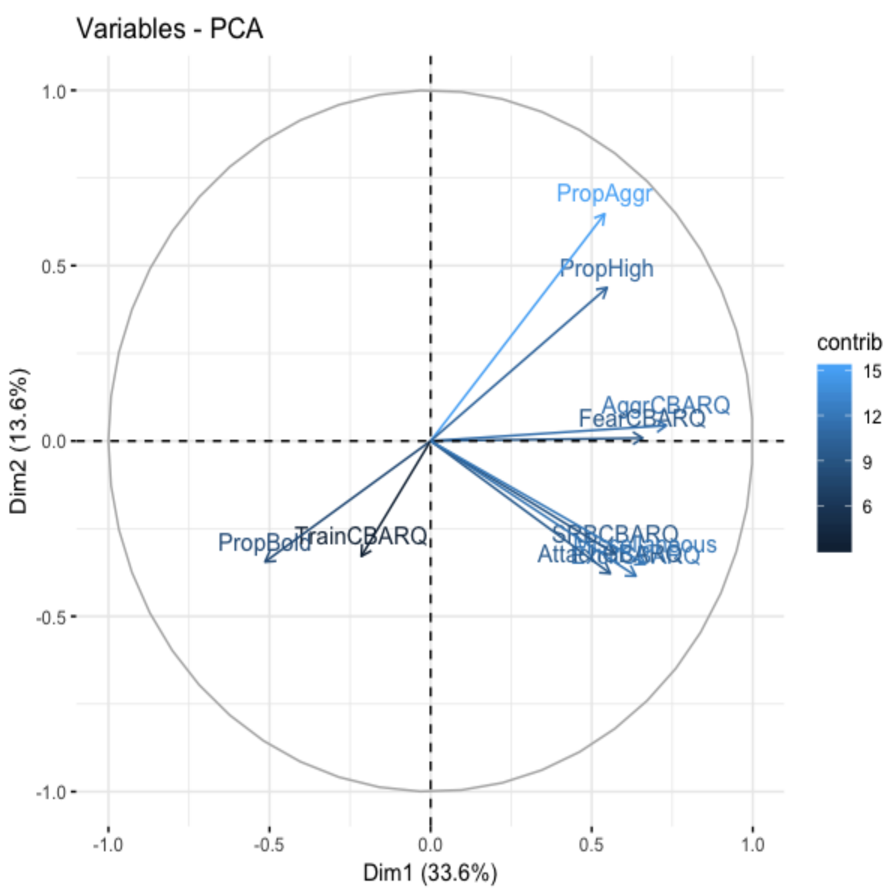
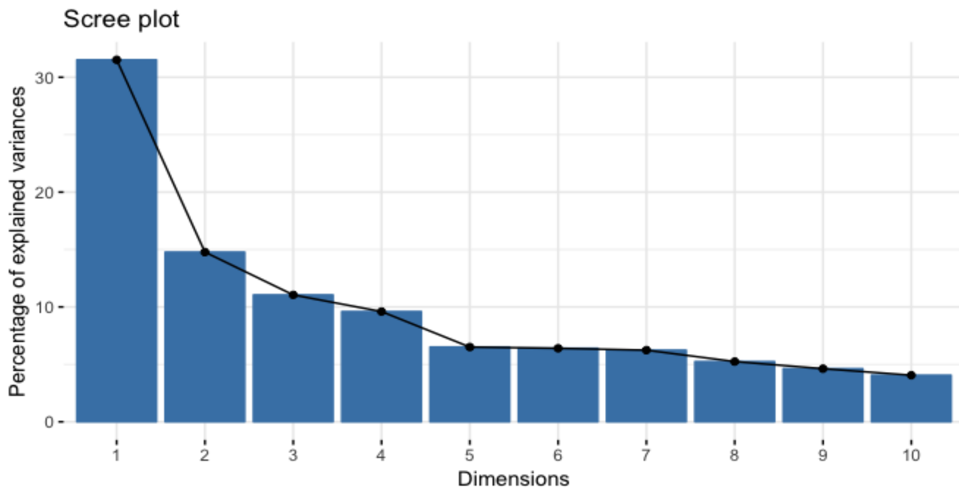


Figure 3. Results of Principle Components Analysis of CBLQ and C-BARQ using dog-directed subscale score. a) Scree plot of the PCA indicating that 2 components are sufficient to explain the variance in scores, and b) Plot of the C-BARQ and BLQ loadings on the two components.

a)



3b.

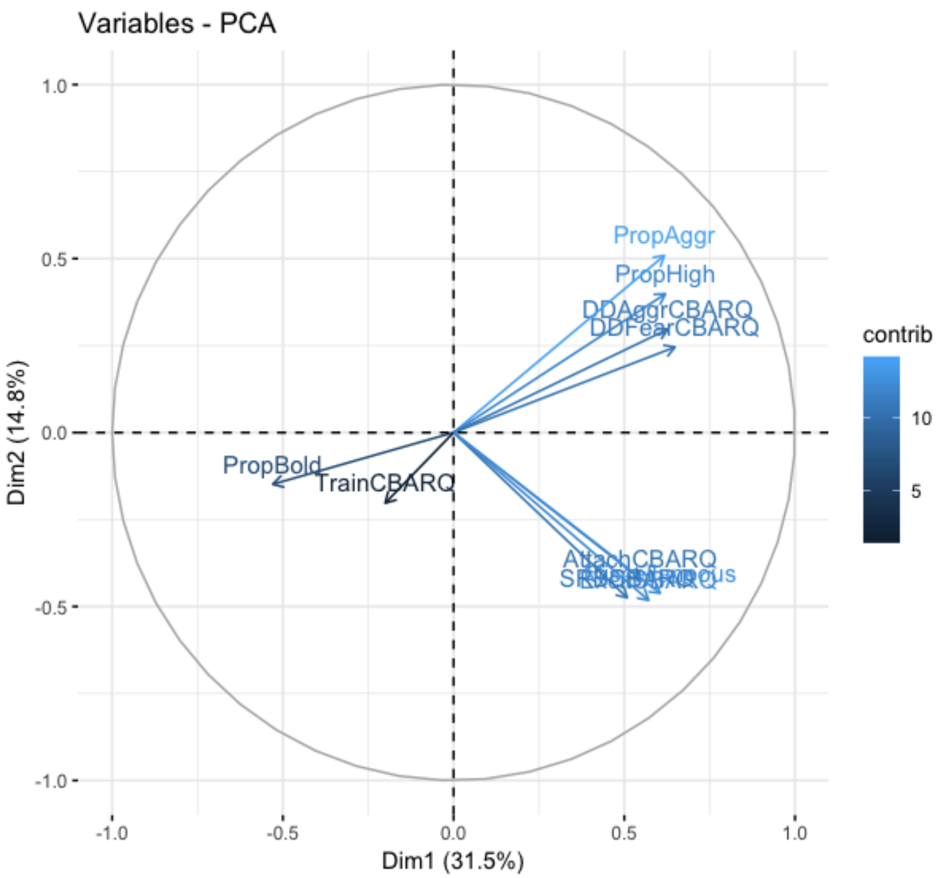


Table 5. Inter-rater reliability of CBLQ dimensions from video recordings.

Stimulus	**Dimension**	**Rater 1 M (SD)**	**Rater 2 M (SD)**	**r**	**p**
Control	Bold	2.042 (0.878)	**2.111** (0.587)	0.637	0.001
	Aroused	2.806 (1.081)	**3.500** (0.702)	0.707	0.000
	Aggressive	**2.931** (0.354)	2.528 (0.519)	0.471	0.020
Novel	Bold	1.861 (0.895)	**2.042** (0.559)	0.514	0.010
	Aroused	2.736 (1.063)	**3.403** (0.816)	0.812	0.000
	Aggressive	**2.778** (0.363)	2.514 (0.590)	0.511	0.011
Model	Bold	**2.153** (0.963)	1.958 (0.751)	0.784	0.000
	Aroused	3.486 (0.868)	**3.861** (0.715)	0.775	0.000
	Aggressive	**2.750** (0.549)	2.569 (0.698)	0.626	0.001
Live	Bold	1.957 (0.949)	**2.087** (0.788)	0.632	0.001
	Aroused	3.493 (0.845)	**3.725** (0.686)	0.676	0.000
	Aggressive	**2.522** (0.658)	2.507 (0.658)	0.604	0.002

Note. The higher score of the two raters in each condition on each dimension is bolded. Rater 2 frequently rated dogs as Bolder and more Aroused, while Rater 1 tended to rate dogs as more Aggressive.

Table 6. Correlation matrix of owner responses on the CBLQ and observer rating when subjects were introduced to a live dog.

	PropBold	**PropAggr**	**PropHigh**	**Bold Rating**	**Aggressive Rating**	**Arousal Rating**
PropBold		$p = 0.36$	$p = 0.68$	$p = 0.54$	$p = 0.12$	$p = 0.69$
PropAggr	-0.19		$p = 0.456$	$p = 0.43$	**$p = 0.0038$****	$p = 0.127$
PropHigh	0.09	0.16		$p = 0.53$	$p = 0.73$	**$p = 0.09$†**
Bold Rating	0.13	0.17	0.13		$p = 0.175$	$p = 0.90$
Aggressive Rating	-0.33	**0.57****	0.07	-0.29		$p = 0.93$
Arousal Rating	0.09	-0.33	**0.36†**	0.03	0.02	

Note. PropBold = proportion of sketches owners selected that were bolder than the matching sketch. PropAggr = proportion of sketches owners selected that were more aggressive than the matching sketch. PropHigh = proportion of sketches owners selected that were higher arousal than the matching sketch. Ratings indicate observer ratings of dogs during introduction to a live dog. * $p < 0.05$, ** $p < 0.01$, † $p < 0.10$

2.7 REFERENCES

Barnard, S., Wells, D. L., Hepper, P. G., & Milligan, A. D. S. (2017). Association between lateral bias and personality traits in the domestic dog (Canis familiaris). *Journal of Comparative Psychology*, *131*(3), 246–256. https://doi.org/10.1037/com0000074

Bear, M.F., Connors, B.W, & Paradiso, M.A. (2001). *Neuroscience: Exploring the brain.* Baltimore: Lippincott, Williams & Wilkins.

Borchelt, P. L. (1983). *Applied Animal Ethology, 10* (1983) 45-61, *10*, 45–61.

Bray, E. E., MacLean, E. L., & Hare, B. A. (2015). Increasing arousal enhances inhibitory control in calm but not excitable dogs. *Animal Cognition*, *18*(6), 1317–1329. https://doi.org/10.1007/s10071-015-0901-1

Brown, C., Brandimonte, M. A., Wickham, L. H. V., Bosco, A., & Schooler, J. W. (2014). When do words hurt? A multiprocess view of the effects of verbalization on visual memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*(5), 1244–1256. https://doi.org/10.1037/a0037222

Carere, C., Van Oers, K., Corr, P. J., Carere, C., Drent, P. J., Privitera, L., … Leea, F. S. (2004). Shy and bold great tits (Parus major): Body temperature and breath rate in response to handling stress. *Physiology and Behavior*, *108*(3), 795–805. https://doi.org/10.1073/PNAS

Corr, P. J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2004.01.005

Corr, P. J. (2008). *The reinforcement sensitivity theory of personality*. (P. J. Corr, Ed.), *The Reinforcement Sensitivity Theory of Personality*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511819384

Demirbas, Y. S., Ozturk, H., Emre, B., Kockaya, M., Ozvardar, T., & Scott, A. (2016). Adults' ability to interpret canine body language during a dog–child interaction. *Anthrozoos*, *29*(4), 581–596. https://doi.org/10.1080/08927936.2016.1228750

Diesel, G., Brodbelt, D., & Pfeiffer, D. U. (2008). Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the UK. *Applied Animal Behaviour Science*, *115*(3–4), 171–181. https://doi.org/10.1016/j.applanim.2008.05.005

Dodman, N. H., Brown, D. C., & Serpell, J. A. (2018). Associations between owner personality and psychological status and the prevalence of canine behavior problems. *Plos One*, *13*(2), e0192846. https://doi.org/10.1371/journal.pone.0192846

Döring, D., Roscher, A., Scheipl, F., Küchenhoff, H., & Erhard, M. H. (2009). Fear-related behaviour of dogs in veterinary practice. *Veterinary Journal*, *182*(1), 38–43. https://doi.org/10.1016/j.tvjl.2008.05.006

Duffy, D. L., Hsu, Y., & Serpell, J. A. (2008). Breed differences in canine aggression. *Applied Animal Behaviour Science*, *114*(3–4), 441–460. https://doi.org/10.1016/j.applanim.2008.04.006

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, *66*(3), 183–201. https://doi.org/10.1037/h0047707

Eysenck, M. W. (1976). Arousal, learning, and memory. *Psychological Bulletin*, *83*(3), 389–404. https://doi.org/10.1037/0033-2909.83.3.389

Fidler, A. E., van Oers, K., Drent, P. J., Kuhn, S., Mueller, J. C., & Kempenaers, B. (2007). Drd4 gene polymorphisms are associated with personality variation in a passerine bird. *Proceedings. Biological Sciences / The Royal Society*, *274*(1619), 1685–1691. https://doi.org/10.1098/rspb.2007.0337

Hsu, Y., & 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*(9), 1293–1300. https://doi.org/10.2460/javma.2003.223.1293

Hsu, Y., & Sun, L. (2010). Factors associated with aggressive responses in pet dogs. *Applied Animal Behaviour Science*, *123*(3–4), 108–123. https://doi.org/10.1016/j.applanim.2010.01.013

Liinamo, A. E., van den Berg, L., Leegwater, P. A. J., Schilder, M. B. H., van Arendonk, J. A. M., & van Oost, B. A. (2007). Genetic variation in aggression-related traits in Golden Retriever dogs. *Applied Animal Behaviour Science*, *104*(1–2), 95–106. https://doi.org/10.1016/j.applanim.2006.04.025

McNaughton, N., & Gray, J. A. (2000). Anxiolytic action on the behavioural inhibition system implies multiple types of arousal contribute to anxiety. *Journal of Affective Disorders*, *61*(3), 161–176. https://doi.org/10.1016/S0165-0327(00)00344-X

Morton, F. B., Lee, P. C., & Buchanan-Smith, H. M. (2013). Taking personality selection bias seriously in animal cognition research: A case study in capuchin monkeys (Sapajus apella). *Animal Cognition*, *16*(4), 677–684. https://doi.org/10.1007/s10071-013-0603-5

Panksepp, J. (1998). *Affective neuroscience: the foundations of human and animal*. *New York: Oxford University Press*.

Plomin, R. (2013). The role of inheritance in behavior. *The Science of Mental Health: Volume 7: Personality and Personality Disorder*, *248*(4952), 229–234.

Rogers, R. D., Lancaster, M., Wakeley, J., & Bhagwagar, Z. (2004). Effects of beta-adrenoceptor blockade on components of human decision-making. *Psychopharmacology*, *172*(2), 157–164. https://doi.org/10.1007/s00213-003-1641-5

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Shi, Z., Hospedales, T. M., & Xiang, T. (2015). Transferring a semantic representation for person re-identification and search. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June-2015*, 4184–4193. https://doi.org/10.1109/CVPR.2015.7299046

Siniscalchi, M., d'Ingeo, S., Minunno, M., & Quaranta, A. (2018). Communication in dogs. *Animals*, *8*(8). https://doi.org/10.3390/ani8080131

Sinn, D. L., Apiolaza, L. A., & Moltschaniwskyj, N. A. (2006). Heritability and fitness-related consequences of squid personality traits. *Journal of Evolutionary Biology*, *19*(5), 1437–1447. https://doi.org/10.1111/j.1420-9101.2006.01136.x

Sinn, D. L., Moltschaniwskyj, N. A., Wapstra, E., & Dall, S. R. X. (2010). Are behavioral syndromes invariant? Spatiotemporal variation in shy/bold behavior in squid. *Behavioral Ecology and Sociobiology*, *64*(4), 693–702. https://doi.org/10.1007/s00265-009-0887-2

Starling, M. J., Branson, N., Thomson, P. C., & McGreevy, P. D. (2013). "Boldness" in the domestic dog differs among breeds and breed groups. *Behavioural Processes*, *97*, 53–62. https://doi.org/10.1016/j.beproc.2013.04.008

Stelmack, R. M. (n.d.). Biological bases of extraversion psychophysiological evidence. *Journal of Personality*, *58*(1), 293-311.

Strandberg, E., Jacobsson, J., & Saetre, P. (2005). Direct genetic, maternal and litter effects on behaviour in German shepherd dogs in Sweden. *Livestock Production Science*, *93*(1), 33–42. https://doi.org/10.1016/j.livprodsci.2004.11.004

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*(1–2), 103–128. https://doi.org/10.1016/j.applanim.2004.08.030
Svartberg, K., & Forkman, B. (2002). Personality traits in the domestic dog (Canis familiaris). *Applied Animal Behaviour Science*, *79*(2), 133–155. https://doi.org/10.1016/S0168-1591(02)00121-1
Tami, G., & Gallagher, A. (2009). Description of the behaviour of domestic dog (Canis familiaris) by experienced and inexperienced people. *Applied Animal Behaviour Science*, *120*(3–4), 159–169. https://doi.org/10.1016/j.applanim.2009.06.009
Walther, D. B., Chai, B., Caddigan, E., Beck, D. M., & Fei-Fei, L. (2011). Simple line drawings suffice for functional MRI decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, *108*(23), 9661–9666. https://doi.org/10.1073/PNAS
Wan, M., Bolger, N., & Champagne, F. A. (2012). Human perception of fear in dogs varies according to experience with dogs. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0051775
Wiener, P., & Haskell, M. J. (2016). Use of questionnaire-based data to assess dog personality. *Journal of Veterinary Behavior: Clinical Applications and Research*, *16*, 81–85. https://doi.org/10.1016/j.jveb.2016.10.007
Wilsson, E., & Sundgren, P. E. (1997). The use of a behaviour test for the selection of dogs for service and breeding, I: Method of testing and evaluating test results in the adult dog, demands on different kinds of service dogs, sex and breed differences. *Applied Animal Behaviour Science*, *53*(4), 279–295. https://doi.org/10.1016/S0168-1591(96)01174-4

Chapter 3. EVALUATING THE VALIDITY OF A MODEL DOG TO ASSESS SOCIABILITY TOWARDS CONSPECIFICS IN THE DOMESTIC DOG *(CANIS FAMILIARIS)*.

Carolyn J. Loyer[1], Kelsey Bezaire[2], James C. Ha[1], Robin L. Foster[1,3]

[1]University of Washington, Seattle, Washington, USA

[2]University of Edinburgh, Edinburgh, Scotland, UK

[3]University of Puget Sound, Tacoma, Washington, USA

Abstract

In companion dogs, aggression towards conspecifics is a common behavior challenge. Assessing whether a dog is aggressive towards other dogs can be logistically challenging and potentially dangerous to humans and to other dogs involved in the assessment. One way to mitigate these issues is to use a model dog in place of a live dog for behavior assessments, however previous research has had inconsistent results when investigating the validity of model dogs for this purpose (Barnard, Siracusa, Reisner, Valsecchi, & Serpell, 2012; Shabelansky, Dowling-Guyer, Quist, D'Arpino, & McCobb, 2015). In the present study, we presented 25 companion dogs with four experimental conditions: a control condition with nothing present, a novel object, a stuffed model dog, and a live dog. We coded dogs' behavior in each condition and used hierarchical cluster analysis to group behaviors for further analysis. Owners completed the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) to gain information about the dogs' behavior in their daily lives. We used Akaike Information Criterion to evaluate models using stimulus condition and C-BARQ scores to predict each behavior cluster. For each

of the behavior clusters evaluated, dogs treated the model dog either similarly to the live dog, or roughly intermediate between the live dog and the novel object. Owner reports of dogs' excitability and their degree of dog-directed aggression (DDA) interacted with stimulus condition to predict whether they treated the model more like a live dog, or more like a novel object. Our results suggest that the inconsistent outcomes of previous studies examining this topic are the result of individual variation in dogs' interpretation of model dogs as conspecifics vs. novel objects. The relationship between DDA and dogs' tendency to treat the model dog more as a conspecific than as a novel object indicates that rather than eliminating the use of model dogs, assessors should add the presentation of a novel object to gain more information about the dogs they are assessing.

3.1 Introduction

Aggression is one of the most serious behavior challenges a dog may present in terms of its potential impact on human and canine welfare. In addition to the possibility for physical harm, agonistic interactions can be costly when considering medical expenses, liability insurance, and possible legal fees. Given the detrimental effect a dog's aggression can have on its owners and society at large, the ability to predict the risk of aggressive behavior prior to rehoming a dog would be financially and ethically advantageous. Past behavior may be predictive of future behavior in similar contexts, but a dog's previous behavior is often unknown if it is found as a stray, or is surrendered or rehomed with an incomplete history from the previous owner.

Multiple behavioral assessments have been developed to identify behavior problems in dogs with unknown or incomplete behavioral histories. The ability of these assessments to predict a dog's behavior in an adoptive or foster home, however, varies widely among those that have been tested (Bennett, Litster, Weng, Walker, & Luescher, 2012; Mornement, Coleman, Toukhsati, & Bennett, 2014, 2015; Valsecchi, Barnard, Stefanini, & Normando, 2011). For example, in one study, nearly half (45%) of the dogs identified as food aggressive in the shelter did not display the behavior in their adoptive homes, while 22% of dogs that did not display food aggression in the shelter did display it in their new homes (Marder, Shabelansky, Patronek, Dowling-Guyer, & D'Arpino, 2013).

Other research investigating the efficacy of behavior assessments has made use of dogs already in stable homes to compare the outcome of these assessments to behavior observed by the dogs' owners. Bennett, Litster, Weng, Walker, and Luescher (2012) compared two assessments commonly used in the United States; the Safety Assessment for Evaluating Rehoming™ (SAFER™, developed by Emily Weiss) and a modified version of the Assess-a-Pet (developed by Sue Sternberg) in their ability to accurately discriminate dogs with a history of aggression from those without. Both tests were found to be marginally effective at identifying dogs whose owners reported a history of aggression via a previously validated behavior questionnaire, the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) (Hsu & Serpell, 2003). Both tests, however, had relatively high error rates in terms of false positives (classifying a dog without a history of aggression as aggressive) as well as false negatives (classifying a dog with a history of aggression as non-aggressive), ranging between 50% (false positives using the SAFER) and 27% (false negatives using the mAAP). This low predictive validity presents a public health concern when dogs' tendency for aggression is missed (Christensen, Scarlett, Campagna, & Houpt, 2007). False positives constitute an animal welfare concern when dogs are labeled as aggressive and potentially euthanized on the basis of an assessment that poorly predicts behavior in the home environment. The present study contributes to the question of whether the equivocal predictive validity in these assessments is due to the (uniquely stressful) shelter environment, or due to the quality of the assessment itself. Behavior assessments used by shelters and rescues are designed to evaluate behavior in a wide range of circumstances that might impact the safety and success of a dog's placement in an adoptive home by recreating situations a dog might encounter in a home environment.

Perhaps one of the most challenging components of behavior assessments is measuring a dog's sociability towards other dogs, in large part due to the logistical issues involved. Ideally, the dog to be assessed would be exposed to a variety of other dogs with known behavioral histories, and its patterns of behavior recorded. This presents two major practical challenges: 1) acquiring and/or maintaining an available group of dogs with known degrees of sociability to act as "helper dogs" during assessments, and 2) safely introducing the dog to be tested to the helper dog(s).

Often, shelters have few dogs to act as helper dogs for sociability assessments. In the United States, a primary goal of many shelters and rescue organizations is to move dogs through their system to adoption (or euthanasia) as quickly as possible. As a result, successful organizations have a rotating collection of dogs to assess, and very few dogs that can act as a helper dog for those assessments. Even if helper dogs are available, additional problems arise: How many helper dogs should a dog be introduced to for an adequate assessment of sociability? What variety of age, sex, and behavioral phenotypes of helper dogs should be included? To what degree should assessments take into consideration order and carryover effects in terms of exposure to other dogs? Should an agonistic interaction occur, is there an appropriate period to allow arousal to decrease before introducing a new helper dog? To our knowledge, many shelters do not have specific protocols in place to address these issues in their assessments, leading to minimal consistency in assessments conducted within a shelters, much less between shelters.

As a solution to the practical and safety challenges of using live helper dogs to assess aggression, shelters, rescue organizations, and canine behavior consultants often use a fake plastic or stuffed model dog (“model dog” hereafter) in lieu of or prior to exposing a live helper dog to the risk of interacting with a dog whose social history is unknown (personal observation via communication with industry professionals). Research examining the validity of using a model dog to assess canine aggression has, however, been inconclusive. Barnard, Siracusa, Reisner, Valsecchi, and Serpell (2012) found that dogs classified as "dog aggressive" via owner reports also showed more fearful and aggressive responses towards a model dog than control subjects with no history of aggression. In a more direct test, Shabelansky, Dowling-Guyer, Quist, D’Arpino, and McCobb (2015) examined the response of shelter dogs to model versus live dogs, and found that the degree of agreement in aggressive behavior between the two conditions was limited: Of the 12 dogs (out of 45 total dogs assessed) that displayed behaviors classified as “aggressive” towards the fake dog, only 3 showed aggressive behavior towards the live dog. There was a greater degree of consistency in “fearful” behaviors between the two conditions, and the highest degree of agreement with “friendly” behaviors, suggesting that validity varies based on the types of behaviors assessed. The authors also noted that the fake dog elicited both more fearful and aggressive responses than the real dog, although the difference was not statistically significant.

The use of a model dog in behavior assessments may result in an inaccurate and inflated picture of the dog’s risk of conspecific social aggression. In order for behavior assessments using model dogs to be useful, we must determine which behaviors are most consistent between model and live dog introductions. We must also understand whether the mixed results of model dog assessments are an artefact of the context (e.g. physical environment, heightened stress induced by a shelter environment, etc.), or a limitation of using a model dog in and of itself. Barnard et al. (2012) found that a novel object elicited a fear response in nearly all canine subjects that had previously displayed aggression towards children or other dogs, compared to less than one third of control subjects with no history of aggression. In Shabelansky et al.'s (2015) study, more dogs were aggressive towards a model dog than a real dog (12 dogs and 8 dogs showed aggression towards the model and real dogs, respectively), and the degree of aggression was more severe towards the model dog. Combined, these findings suggest that a model dog may elicit a greater negative response from dogs than would be elicited by a live dog. Two possible explanations are that the model dog is perceived as 1) a novel object (eliciting increased fear-based aggression particularly in neophobic dogs), or 2) as a threatening and socially inappropriate social stimulus.

Do dogs with more or less sociability respond similarly to model dogs? Previous research has examined the reliability of dog-dog sociability assessments conducted in shelters, and separate studies have looked at the degree of agreement in dogs' behavior when presented with live versus model dogs, but to our knowledge, no peer-reviewed studies to date have compared the responses of dogs with known histories of conspecific sociability or aggression towards live versus model dogs. Further, we know of no published research that has examined whether dogs' behavior towards a model dog is more similar to their response to a live helper dog, or towards an inanimate object, which led us to our first research question. Our second question was whether individual differences in behavior predict dogs' behavior in the experimental set up, particularly how their behavior varied towards each of the stimuli. To address these questions, we used owner reports of dogs' behavior in their daily lives as well as behavioral observations of the dogs when presented with four experimental conditions: a live dog, a stuffed model dog, a novel object, and a control condition with no object present. . Further understanding what factors may explain individual variation in dogs' responses to these stimuli may help explain the moderate consistency and predictive validity found in past work investigating the validity of using model dogs to assess canine behavior.

3.2 METHODS

3.2.1 Subjects

Dog owners were recruited via social media and word of mouth to participate in the study. We did not restrict participants to particular breeds in order to increase the generalizability of our results. We chose to use dogs that were currently in stable homes, as they offered the best opportunity to see if their behavior in our experiment was a good predictor of their behavior in their normal lives. Interested owners were given a brief description of the study requirements and asked to fill out an intake survey gathering information regarding dog demographics (age, breed or mix of breeds, how the dog was acquired, etc.), and history of aggression towards humans and other dogs. We excluded dogs with a history of serious aggression towards humans for safety purposes. In order to participate, owners must have lived with their dogs for at least one year (thus, all dogs had to be at least one year of age), and the dogs must not have any significant mobility or physical limitations. Dog-owner pairs that met these requirements and were available during the data collection period were given detailed participation instructions and a consent form in advance. Twenty-six dog-owner pairs participated. One subject's data was excluded because the dog did not meet the age criteria, resulting in a total sample size of N = 25 (see Table 1 for a complete list of participating dogs).

3.2.2 Ethical statement

Owners were provided with detailed instructions and a consent form via email in advance of their participation. They were instructed that they could opt out of participation at any time. This protocol was approved by the University of Washington Institutional Animal Care and Use Committee, protocol # 201700112.

3.2.3 Experimental set up

All sessions took place at an indoor dog training facility in Seattle, WA. The room was 74’ x 42’, with rubber mat flooring throughout for traction. Mobile chain link fences 5’ H x 10’ L were set perpendicular to the walls at six- to seven-foot intervals on the two longer sides of the room. For the purposes of the experiment, we further segmented the space using one of the chain link fences and two X-pens which were clipped to the chain link. This allowed for physical separation of the subject dogs from the location where the stimuli were presented, while still allowing visual, olfactory, and auditory contact. Three markers were placed in a line at distances of 32’, 20’, and 8’, respectively, from the fence separating the subjects from the stimulus. A fourth marker (the proximal, or contact, marker) was placed inside the stimulus presentation area at a distance of 3’ from the stimulus. In a back corner of the room, we created a “blind” area by covering one of the chain-link fences with a curtain. This prevented the subjects from viewing stimuli as they were brought into the room. All sessions were recorded from two angles using a Canon Vixia HFR600 and a Sony Handycam HDR-CX240 for later behavioral coding. See **Figure 1** for the layout of the study site.

3.2.4 Stimuli

Live dog: The same stimulus dog was used in the live dog condition for all subjects: An 8.5 yo spayed female Labrador retriever/standard poodle cross with a black and gray wire-haired coat, measuring 26" at the shoulder and 29" from the base of the neck to the base of the tail, and belonging to CL. The stimulus dog had extensive experience as a "neutral" or stimulus dog for other dogs in obedience training and in rehabilitation of dog-dog reactivity. To reduce variability in stimulus dog behavior, she was cued to stand or sit in the designated location, facing perpendicular to the approach of the subjects, and was offered treats to keep her attention on the handler (CL). If the stimulus dog turned her gaze towards the subject dog, she was lured with a treat to return her attention to the handler. At no point during the trials did she bark or otherwise vocalize. For all other conditions, CL kneeled in the same position behind the stimulus and mimed offering treats.

Model dog: The model dog used was a stuffed Melissa & Doug™ gray/silver Labrador retriever in a standing position, measuring 19" at the shoulder and 22.5" from the base of the neck to the base of the tail. The model dog was positioned so that its body was perpendicular to the subjects' angle of approach, while its head—positioned at a fixed angle toward the left—was turned towards the subjects.

Novel object: For the novel object condition, a stool was laid on its side and covered with a blue fitted sheet to create a stimulus roughly 13" H x 13" W x 30" L.

The color of the sheet was selected to be within the range of wavelengths for which dogs can discriminate colors (Neitz, Geist, & Jacobs, 1989). In order to control for potential scent cues, the model dog and the sheet used in the novel object condition were both laid on the dog bed used by the live dog for approximately one hour prior to testing to allow them to absorb the scent of the live dog.

Control/null condition: In the control condition, CL kneeled in the same position and location as for the other stimuli and offered treats to the space in front of her. The control condition provided a baseline of each individual dog's behavior in the testing environment, since it is possible that their behavior may differ in general, as well as in a novel location.

Presentation order of the stimuli was completely counterbalanced to control for order effects, and the order each dog received was determined randomly. One owner elected not to have their dog (Oliver) participate in the live dog condition due to extreme fear. We included data for this dog, and added an additional subject to maintain complete counterbalancing using all conditions.

3.2.5 Procedure

Upon arrival, owners were asked to leave any dog treats or toys with the experimenters, who placed the items out of sight. Each owner attached a provided, 5' nylon flat leash (.75" wide) with a padded handle to their dog's collar or harness. For smaller dogs, a lighter, narrower leash (.5") of the same length and design was used. Once the dog's leash had been exchanged, one experimenter (RF) invited owners to let their dog explore the room for five minutes. Dogs remained on leash during this time, and RF and the owner followed the dog around the room. During this habituation period, RF reviewed the study instructions for the owners and coached owners not to use the leash, verbal commands, or their body language to cue the dogs' behavior during trials. Between trials, while they were returning to and waiting in the blind area, they were allowed to interact with their dog as normal.

Owners and their dogs waited in the blind area after the 5-minute acclimation and between each stimulus condition. CL entered the room carrying the stimulus and placed it in a standardized location, oriented so that the longest side of the stimulus was facing the subjects. For the live and model dog conditions, this meant that the spine was perpendicular to the line of markers and the owner's approach. Across all conditions, two auditory cues were presented before the subject dog left the blind: 1) the jingling of a set of dog tags as CL entered the room, and 2) a bark recorded from the live stimulus dog during a play session with CL in the research space, played over a BOSE SoundLink wireless Bluetooth speaker, model# 415859. After the recording concluded, RF prompted owners to exit the blind and walk to first marker on the floor (the marker closest to the blind, and 32' from the stimulus presentation area). Owners were asked to stop with their feet on the marker for ten seconds, then prompted to walk forward to the next marker.

For all but the live dog condition, after the pause at the 8' marker RF opened the gate between the subject dog and the stimulus and asked owners to walk to the proximal marker. This allowed the dogs to freely interact with the stimulus within the range of their leash. If the dog did not interact with the stimulus within 20 seconds, the trial ended. If the dog did approach the stimulus, they were allowed to continue to investigate and sniff until they had disengaged from the stimulus and kept their attention elsewhere for at least 20 seconds.

To reduce experimenter bias the following controls were set in place:

1. The experimenter acting as the handler for the stimuli (CL) minimized interaction with owners and dogs prior to their participation to reduce the risk of affecting dog behavior during the stimulus presentation.

2. CL was unaware of each dog's history of aggression, though RF had previously screened each owner-dog team. While ideally both experimenters would be completely blind to each subject's history, for safety reasons it was decided that this was inadvisable.
3. During each trial, CL knelt in the same position regardless of the stimulus, and directed her gaze towards the stimulus, not at the subject dog or owner.
4. RF maintained a neutral facial expression and stood behind one of the chain link fences, which was also covered with a curtain to mask any body movement.
5. CL was not aware of the stimulus order until cued by RF to retrieve the next stimulus.

3.2.6 Owner-completed behavior survey

At least two weeks after owners completed the in-person participation, they were asked to complete follow up survey to collect more detailed information regarding their dog's behavior. The survey was created and responses collected using Qualtrics online software (Provo, UT). The survey included questions regarding the frequency of aggressive interactions between subject dogs and conspecifics, the greatest degree of injury inflicted during those interactions (to the subject and to the other dog), the variability of the subject's behavior towards other dogs, the degree to which owners could predict their dog's behavior towards other dogs, and the factors influencing the variation in their dog's behavior. It also included a previously validated questionnaire for assessing canine behavior, the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) (Hsu & Serpell, 2003). The C-BARQ is a previously validated questionnaire that identifies a number of potential individual differences in canine behavior, particularly potential behavior problems that owners may encounter (e.g. severity of fearful behavior, aggression, and dogs' tendency to overreact to changes in their environment) (see

Table 2 for a complete list of C-BARQ scales and subscales used in this study). This questionnaire is frequently used in behavior research (Barnard, Wells, Milligan, Arnott, & Hepper, 2018; Hsu & Sun, 2010; MacLean et al., 2017), and offers a relatively efficient method of quantifying behavior in dogs' daily lives. We received complete survey responses from twenty-three of the twenty-five participating dog owners, partial data from one owner, and no response from one owner.

3.2.7 Ethogram development & behavior coding

We conducted a literature review in order to develop an ethogram of dog behaviors that included behaviors typically associated with both positively- and negatively-valenced emotional states (Barnard et al., 2012; Beerda, Schilder, Van Hooff, & De Vries, 1997; Beerda, Schilder, Van Hooff, De Vries, & Mol, 1998; Borg, Schilder, Vinke, & Vries, 2015; Cannas et al., 2014; Deldalle & Gaunet, 2014; Döring, Roscher, Scheipl, Küchenhoff, & Erhard, 2009; Ogata, Kikusui, Takeuchi, & Mori, 2006; Ottenheimer Carrier, Cyr, Anderson, & Walsh, 2013; Shabelansky et al., 2015; Sommerville, O'Connor, & Asher, 2017). The behaviors were selected with emphasis on fear-, anxiety-, and play-associated behaviors. We further scored the average duration each dog spent with their gaze directed towards the stimulus, and dogs' body positions based on the average weight distribution, ear, tail, and mouth positions. Gaze duration is a potential proxy for level of interest in the stimulus, which we expected to vary with each condition based on whether the stimulus socially relevant and dogs' motivation to engage in social interactions. We tested our ethograms using pilot videos of the experiment to rule out behaviors too vague or difficult to observe. See Tables 3 and 4 for a complete list and description of the behaviors included in the final analyses.

Subject behavior was coded from videos taken during each session. Behavior at a distance (when owners stood on the markers 32', 20', and 8' from the fence separating subjects from the stimulus) was coded separately from behavior on contact with the stimulus (when owners stood on the proximal marker, and subjects were allowed to interact with the stimulus). On contact, we recorded the duration of time dogs spent engaging with the owner, the stimulus, the experimenters, or the environment. Sniffing and engagement was categorized as "Other," "Owner-directed", "Social" (towards the stimulus' face or urogenital region on the model dog, and in corresponding locations on the novel object) or "Non-social" (directed towards the stimulus, but not in a socially-relevant location). See Table 5 for behavior coding categories and operational definitions by stimulus on contact.

3.2.8 Statistical analyses

Differences between stimuli in average gaze duration, average time spent sniffing, and average proportion of sniffing focused on socially relevant locations (social sniffing) were compared using repeated measures ANOVAs. When indicated by the omnibus test, post-hoc pair-wise comparisons were conducted using Fisher's LSD t-tests. We used Pearson's product moment correlation to examine the relationship between the amount of time dogs spent sniffing the model dog and novel object on contact, and between the amount of time spent sniffing socially relevant areas on each stimulus.

To assess whether the behaviors coded at the three distal markers were correlated, we used hierarchical clustering using the ICLUST package in R (version 3.6.0) to identify behavior factors. Our criterion for combining behaviors into a factor was that the correlation between behaviors was ≥ 0.5. In order to include a higher order correlation, Cronbach's alpha had to be greater than .6 with a correlation of at least .5 with any lower order clusters. To determine whether owner reports of dogs' behavior in their daily lives predicted dogs' behavior in our experimental set up, we used Akaike Information Criterion with a correction for a small sample size (AICc). C-BARQ scores and experimental condition were entered fixed effects, with Dog ID as a random effect and the selected factor scores as the outcome variables. Because one main goal of this study was to identify behavior differences that predict whether dogs will respond to the model as if it is a live dog versus a novel object, when the top model selected via AICc did not include any CBARQ scores as predictors, we then examined the next best model. If the next model had a difference in AICc of less than three, we used the MuMIn package in RStudio to obtain model averaged coefficients and p-values to determine best predictors of the behavioral factors.

3.3 RESULTS

Average age (± SD) of the subjects was 4.82 years (± 2.43yrs), and dogs had been with their current owners for an average of 3.83 years (± 2.25yrs). Owners had acquired their dogs in a variety of ways, including from a breeder when the dog was a young puppy (n = 6), from a private rescue organization (n = 8), from a public animal shelter (n = 10), or from the previous owner (n = 1). Breed and size of the dogs varied widely, including a range of purebred and mixed-breed dogs, categorized based on owner report (Table 1).

Preliminary analyses revealed that there were no significant or systematic differences across the three distal markers in subjects' average duration spent gazing at the stimulus, in average body positions, or in behaviors coded as present vs. absent. Results presented here represent individual dogs' behavior averaged across the distal markers.

3.3.1 ANOVA of gaze duration at a distance by condition

The average amount of time dogs spent staring at the novel object was similar to that of the control condition, and the average amount of time spent staring at the model and live dog was similar and significantly greater than the other two conditions. A repeated measures ANOVA determined that average gaze duration directed toward the stimulus at a distance varied significantly between stimuli (F (3, 69) = 18.16, $p < .0001$). Post hoc comparisons revealed that there was no significant difference in average gaze duration (± SD) between the control condition (M = 2.91 sec ± 2.07) and the novel object (M = 2.81 sec ± 1.80, p = 1.00), or between the model dog (M = 4.93 sec ± 2.68) and the live dog (M = 5.73 sec ± 2.16, p = .13). Average gaze duration in the control condition was significantly less than in the model dog and live dog conditions (p = .003 and $p < .001$, respectively). Average gaze duration in the novel object condition was significantly less than in the model dog (p = .005) and live dog ($p < .001$) conditions (Figure 2).

3.3.2 Correlations between individual dogs' average duration staring by stimulus

The time dogs spent with their gaze directed toward the model dog was strongly positively correlated with their average gaze duration toward the live dog (r = .729, $p < .0001$). Average gaze duration towards the novel object was significantly positively correlated with the control condition (r = .523, p = .006). Average gaze duration in the control condition trended towards a moderate positive correlation with average gaze duration in both the model (r = .486, p = .012) and live dog conditions (r = .398 p = .049, Table 6).

3.3.3 Comparison of sniffing behavior toward the model dog and novel object

Dogs' sniffing behavior suggested that they were treating the model dog more as a social stimulus than the novel object. Overall time spent sniffing the model dog (M = 12. 92 ± 16.62sec and novel object (M = 12.12sec ± 10.48) did not differ significantly ($t(24) = .265$, $p = .793$, Figure 3), and there was a trend towards a positive correlation between overall time individual dogs spent sniffing the model dog and the novel object ($r = .417$, $p = .034$). For dogs that sniffed both the model dog and the novel object, dogs tended to spend a greater proportion of time sniffing socially relevant areas on the model dog ($M = .76$, $SD = .20$) than on the corresponding areas of the novel object ($M = .53$, $SD = .25$, $t(17) = 2.73$, $p = .014$, Figure 4). Dogs that were attracted to the socially relevant areas on the model dog were not more likely to sniff the corresponding locations on the novel object. There was no significant relationship between the proportion of time a dog spent sniffing socially relevant areas on the model dog and the corresponding areas on the novel object ($r = -.230$, $p = .359$).

3.3.4 Behavior clusters and factor selection.

Hierarchical cluster analysis and our criteria for grouping behaviors resulted in five factors (Figure 5). “Freeze” and “low posture” loaded positively on the first factor. As these behaviors have been found to be elicited by the acute presentation of aversive stimuli (Beerda et al., 1998; Ogata et al., 2006), we labeled this factor “Fear.” “Jump” and “Bark” loaded positively on the second factor, which we named “External Arousal.” Owners of dogs that are reactive towards other dogs commonly complain about these behaviors directed towards other dogs on leash walks, and barking is used as an indicator of excitability on the C-BARQ.

“Average gaze” and “Orient” loaded positively on the third factor, while “Average body angle” and “Explore” loaded negatively, indicating that dogs that spent more time with their gaze and body directed towards the stimulus, or that oriented towards something in the environment more often did not explore environment as frequently as dogs that were less focused on the stimulus. We named this cluster "Stimulus Focus." Increased attention may be triggered by both appetitive and aversive stimuli, a higher score on this factor does not necessarily indicate the emotional valence of the stimulus presented. However, we expected that stimuli that elicit similarly intense emotional responses (positively- or negatively-valenced) would elicit similar scores on this factor.

The fourth factor included positive loadings for "Wag," "Whine," "Pant," "Pace," and "Liplick." As most of these behaviors occur more often when dog are in stressful situations (Beerda et al., 1997, 1998; Ogata et al., 2006) we labeled this factor "Anxiety." The final factor included positive loadings for "Paw at" and "Jump at Owner". However, because each of these behaviors occurred rarely in our sample (five trials out of ninety-nine total) and were not primary behaviors of interest, we excluded this factor from further analysis. "Bowing," which is presumed to be pro-social and positively-valenced, did not occur during our testing.

3.3.5 Model selection to predict behavioral factors.

Stimulus Focus:

Based on AIC, Stimulus Focus was best predicted by the experimental condition (Table 7). The model that included C-BARQ Aggression scores had a Δ of 1.47 higher than the model including only experimental condition, thus we used the model averaged coefficients of the two models obtained using the MuMIn package in R. There was not a significant main effect of C-BARQ aggression score on Stimulus Focus ($z = 0.793$, $p = 0.428$). There were differences in Stimulus Focus by condition. Dogs focused less on the stimulus in the control condition compared to all other conditions ($z = 3.664$, $p = .000248$). Stimulus Focus was significantly higher for the live dog compared to the novel object (*z* = 4.454, *p* = 8.4 x 10^{-6}). Stimulus focus towards the model dog did not differ from the midline between the novel object and the live dog (*z* = 0.225, *p* = 0.82, Figure 7). C-BARQ Aggression score interacted with stimulus condition to predict dogs' tendency to orient toward the stimuli. As Aggression score increased, dogs stimulus focus increased toward the model dog and the live dog (*z* = 1.973, *p* = 0.049), but did not change for the control (*z* = 0.143, *p* = 0.89) and novel object conditions (*z* = 0.734, *p* = 0.46, Figure 8).

External Arousal:

Jumping and barking were best predicted by the model that included C-BARQ Composite Arousal score and its interaction with the stimulus condition (Table 8). There was a positive effect of Composite arousal score on the probability of jumping and barking overall (*z* = 2.623, *p* = 0.008715). There was not a significant interaction between Composite Arousal and the difference in behavior between the novel object and the live dog (*z* = 1.384, *p* = 0.166443), or between the model dog condition and the novel/live conditions (*z* = .596, *p* = .551345) but there was an interaction with the difference in behavior between the control compared to all other conditions. Dogs that were more excitable showed more jumping and barking in general, and showed less variation in their behavior across conditions compared to dogs that scored lower on Composite Arousal (*z* = 3.310, *p* = 0.000933).

Dogs showed greater External Arousal (i.e., they jumped and barked more often) when a stimulus was present than in the control condition(*z* = 3.419, *p* = 0.000628), and in the live condition compared to the novel object (*z* = 2.808 , *p* = 0.00498). Dogs showed more jumping and barking in the model dog condition than in either the live dog or novel object conditions (*z* = 2.403 , *p* = 0.016, Figure 9).

Anxiety:

Anxiety behaviors were best predicted by the model that included only stimulus conditions (Table 9). Dogs were less likely to display these behaviors in the control condition compared to the others ($z = 2.525$, $p = 0.0116$). They were significantly more likely to display these behaviors in the live condition compared to the novel object ($z = 3.543$, $p = 0.000396$), and the likelihood of these behaviors in the model dog condition was intermediate between the novel object and live dog conditions ($z = 0.893$, $p = 0.372$, Figure 10).

Fear:

The probability of dogs' freezing and low posture was best predicted by the null model (Table 10), indicating that the tendency to produce these behaviors was unique to individual dogs and was not dependent on the stimulus presented, nor predicted by CBARQ scores. Closer examination of the data revealed that although Fear behavior occurred sufficiently often to be included in our cluster analysis, these behavior were still very rare in our experiment. Only two of our subjects exhibited either freezing or low posture in more than one condition. One of these subjects displayed Fear behaviors in the control and model dog conditions. The second subject displayed both behaviors in all four conditions, with the most intense response in the control condition. Thus, the tendency to freeze or display low posture was idiosyncratic in our experiment, and our results for this behavior cluster are not interpretable.

3.4 Discussion

Our results provide a potential explanation for the previously mixed findings studying the use of a model dog to assess canine intraspecific sociability. Dogs' behavior towards the model dog was more like their behavior towards a live dog than towards a novel stimulus. In fact, dogs treated the model and the live dog similarly when comparing gaze duration, jumping and barking, and freezing and low posture. Dogs also spent marginally more time sniffing socially relevant areas on the model dog than they did on the corresponding areas of the novel object, although the average time spent sniffing each stimulus did not differ. Each of these results suggest that the model dog elicits behavior more like that directed towards a conspecific than a novel object. However, when it came to behaviors such as lip licking, whining, panting and pacing, as well as orienting their body and gaze towards the stimulus, dogs' behavior towards the model dog was intermediate between that of a novel object and the live dog. This suggests that the model dog is eliciting some, but not all, responses another dog.

There are multiple reasons that these differences might exist. One potential explanation is the difference physical appearance of our live dog compared to the model dog. The live dog was slightly larger than the model dog, and a different color, which may have affected how dogs responded to the two stimuli. Perhaps more importantly, the model dog and live dog showed differences in "behavior". In social interactions, behavior depends on the actions of the social partner. Since the model dog does not generate or respond to social cues, our subjects received minimal social feedback. This likely alters their behavior compared to an interaction with a live dog which reciprocates and responds to social cues. The inability of model dogs to react appropriately is one of the primary arguments against using them in behavioral assessments.

In some cases, dogs' behavior towards the model dog was more extreme than towards the live dog (e.g. Fear). This provides additional support for previous research indicating that a model dog may provoke more negatively-valenced behavior than a real dog might. Depending on the goal of the assessment, this may be beneficial or detrimental. If the goal is to determine how a dog will respond to a supernormal social stimulus (i.e. a dog with questionable social skills), it may be valuable to present them with a model dog that is stiff, unmoving, and stares directly at the subject. This may give some measure of whether a dog is likely to flee or fight, and the degree of aggression they display if it is elicited. However, if the aim of an assessment is to determine whether a dog is likely to be friendly towards “the average dog” that they meet while on a walk with their owners, the model dog’s tendency to elicit more extreme fear behaviors in our study, and more aggression in Shabelansky et al.'s study (2015) may falsely exaggerate dogs’ negatively-valenced behaviors. In shelters with limited space and an excess of animals, this may put dogs at higher risk of euthanasia based on an imperfect assessment.

Additionally, based on our results and those of previous researchers in this area, interpretation of dogs’ behavior towards a model dog should be conducted with some skepticism and as part of a more complete behavior evaluation. This caution is even more important when evaluations are being conducted on recently surrendered or rehomed dogs. Owner reports of dogs’ behavior predicted some, but not all of the behavior factors identified in our experiment. This may be due to the novelty of the context to most of our subjects. If the context (or the novelty) of a behavior assessment has the potential to alter a dog’s behavior sufficiently that it is not predicted by owner reports of behavior, we must be even more careful in interpretations of assessments conducted on dogs in the midst of a stressful transition in the fluid shelter environment.

A further challenge of assessing dog-dog sociability in the shelter is that the validity of the assessment can be compromised if shelter staff and volunteers do not have advanced training in canine behavior, which is relatively common (Mornement, Coleman, Toukhsati, & Bennett, 2010), and which poses a safety hazard. Failure to recognize precursors to agonistic interactions may go unnoticed, increasing the chances of a fight and therefore the risk of injury to both dogs and humans during assessment. When assessing dogs’ behavior towards another dog, Diesel, Brodbelt, and Pfeiffer, (2008) found low inter-rater agreement in ratings of “fear-aggressive,” “indifferent,” or “pushy-aggressive” responses between shelter staff. Agreement increased slightly for most categories of behavior when the sample was restricted to shelter staff that either had more than 8 years of experience working with and assessing dogs, or who had formal training in canine behavior. This suggests that in order to improve the validity of an assessment, it should be conducted and scored by experienced or well-trained individuals.

The results of this study have implications for rehabilitation of dog-directed aggression. It is possible that improving dogs’ aggression may depend more on our ability to help them discriminate and respond appropriately to social stimuli than to countercondition a presumed fear response. The interaction between dogs’ DDA C-BARQ scores and their fear-based behaviors based on the stimulus was interesting and unexpected. Aggression towards other dogs is commonly thought to stem from fear, but in our study we found that freezing and low posture when presented with the live dog and model dog were less common in dogs whose owners reported higher levels of dog-directed aggression, and that these dogs tended to treat the model dog more as they did the live dog compared to dogs who were scored lower on DDA. This suggests that for some dogs, fear is not the motivating emotion behind their reactivity towards other dogs, and that their behavior towards other dogs is more generalized compared to dogs who are less reactive. This has important implications for addressing dog-directed aggression and reactivity.

The main limitation of this research is our sample size, exacerbated by the rarity of the behaviors of interest. The infrequency of the behaviors measured means that a larger sample size is necessary to have the statistical power to detect individual differences that predict their occurrence. This was somewhat mediated by increasing the amount of time dogs spent at a distance from the stimuli, and by averaging behaviors across the three distal markers. However, given the complexity of the analyses required to determine inter-stimulus differences as well as inter-subject differences, we would like to see this work replicated with a larger subject pool, preferably with greater representation of dogs on the extremes of the behavior spectrum. Our results indicate that dogs' tendency to display aggressive behavior, as well as their excitability in novel situations may affect whether they treat model dogs as social stimuli or as novel objects. Future research may benefit from focusing on comparing dogs at both ends of these behavioral dimensions.

One additional challenge to interpreting our results is the way in which subjects were introduced to the stimuli. The effect of the ten second pause at each marker may have affected dogs' responses. At each distance, they had an opportunity to observe the stimulus with minimal interaction with or direction from their owner. This prolonged approach may have improved dogs' ability to discriminate the model dog from a live dog. It is also possible that it could have exaggerated anticipatory behavior, either pro-social or agonistic. While none of the dogs in our sample snapped, snarled, or growled loudly enough to be coded in our ethogram, multiple dogs pulled on their leash towards the stimulus, and some performed what may have been frustration-related behavior (e.g. playing tug of war on the leash) during their approach. In the outside world, dog-dog interactions are sometimes more fluid, as when dog owners walk down the sidewalk towards each other, or when dogs encounter each other off-leash. Our experimental design was arranged to examine whether dogs' behavior towards the stimuli would change as they approached, perhaps due to improved multi-modal information (e.g. detecting a difference in smell between the model dog and the live dog, or improved ability to notice social cues or lack thereof), and to standardize the approach speed of the owners. However, preliminary analyses detected no differences across the distances, thus, we collapsed responses to improve our statistical power. We suggest that for future work, approaches are done without pause, and perhaps use visual markers to distinguish behavior at a distance compared to in proximity. In this way, allowing the dog to lead the approach speed would give additional information in terms of latency to make contact with the stimulus, which may also be an important predictor of sociability or aggression.

Our work has important implications for the validity of using a model dog to assess canine aggression towards conspecifics. Understanding that dog-directed aggression may influence whether dogs treat the model dog as a live dog versus a novel object allows us to better understand the information provided when we use a model dog for behavior assessments. It may be beneficial to present dogs with novel objects as well as a model dog *prior* to conducting a behavior assessment with a live dog in order to better understand each dog's behavior. Our results indicate that in general, dogs' behavior toward a model dog is more similar to that directed toward a live dog than a novel, inanimate object, but that this varies based on the type of behavior measures. Importantly, the model dog may elicit more negatively-valenced responses than a live dog. This suggests that dogs who respond with fear or aggression toward a model dog should not be immediately categorized as dog-aggressive, but may indicate a need for additional care taken in their introductions to conspecifics during further evaluation.

3.5 TABLES AND FIGURES

Table 1. Name, age, and breed of companion dogs that participated in the research.

Dog Name	Dog Age (years)	Breed (reported by owner)	Weight (lbs.)
Angel	7.5	Mini Australian Shepherd	25
Asia	5.5	Doberman pinscher	70
Bambu	5.5	Hound/Labrador Retriever mix	40
Bo	5	Chihuahua	7
Cooper (1)	2	Vizsla	52
Cooper (2)	1.5	Pit bull/German Shorthaired Pointer mix	55
Dale	4	Kelpie/Australian Cattle dog mix	30
Dexter	1	American Pitbull terrier	60
Franklin	6	Dachshund/Fox terrier mix	17
Johnny Cash	5.5	Chihuahua/Rat terrier mix	12
Jorgen	5.5	Plott Hound	65
Kit	6	Papillon	14
Lola	3	Shepherd/Hound mix	45
Nior	8	Doberman pinscher	95
Oliver (1)	7	Shih Tzu/Poodle mix	12.5
Oliver (2)	4	Longhaired Chihuahua	10
Rager	3.5	Smooth-coated Border Collie	40
Ramsey	4.5	Chow/Shepherd mix	55
Rollie	10.5	Rottweiler/Beagle mix	65
Sadie	5	Pembroke Welsh Corgi	27
Sirius Black	10	Chow mix	55
Sly	4.5	Parsons Russell Terrier mix	12
Tally	1.5	Border Terrier	17
Truffle	1.5	Pit bull-type mix	65
Zoe	3	Chihuahua/Terrier mix	11

Figure 1. Diagram of experimental set up. Owners waited with their dogs in the area labeled "Waiting" between trials, while one experimenter brought the stimuli into the room and placed them in the area marked "Stimulus." A second experimenter cued the owner to approach the stimulus, pausing for 10 seconds at each X marked in the diagram. For all conditions except for the live dog, the trial ended after the subject had an opportunity to interact with the stimulus while the owner stood on the X directly next to the stimulus. For the live dog condition, the trials ended after the third marker: subjects did not interact with the live dog. Two cameras recorded the trials for later behavior coding.

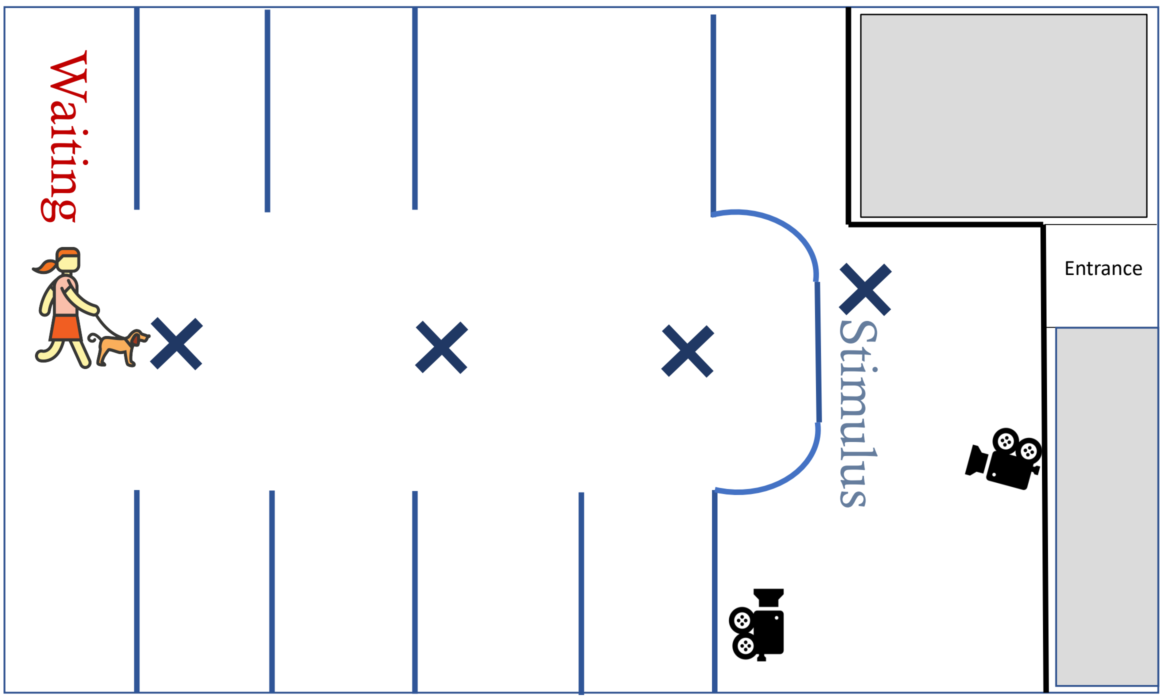


Table 2. C-BARQ scales, subscales, and descriptions

Scale	Description
Trainability	Responsiveness to owner cues, sensitivity to potential rewards and aversives. Higher score = More “trainable”
Aggression	Severity of aggression displayed in a variety of situations, e.g when verbally or physically corrected, during grooming, etc. Higher score = more severe aggression
Dog-Directed Aggression (DDA)	Severity of aggression displayed during interactions with other dogs. Higher score = more severe aggression
Fear and Anxiety	Severity of fear displayed in a variety of situations, e.g. when hearing a vacuum cleaner or fireworks, or when an unfamiliar human approaches. Higher score = more severe fear
Dog-Directed Fear (DDF)	Severity of fear displayed during interactions with other dogs. Higher score = more severe fear
Separation-related behavior (SRB)	Frequency of particular signs of distress are displayed when the dog is left alone. Higher score = more frequent SRB
Excitability	Degree of reaction to a variety of situations, e.g. when the doorbell rings or before a car ride. Higher score = more excitable
Attachment and Attention-seeking	Frequency with which dog follows or solicits attention from a particular human household member. Higher score = more attachment
Miscellaneous	Frequency of other potential behavior concerns, e.g. eating feces, excessive licking. Higher score = more frequent behaviors.
Composite Arousal	A combination of Excitability and questions from the Miscellaneous scale related to arousal, e.g. a dog’s tendency to be “hyperactive, restless, have trouble settling down.” This score was calculated for the purposes of this study and is not a part of the original C-BARQ.

Table 3. Behaviors coded on a scale at each distal marker

Behavior	**Description**
Gaze duration (seconds)	The total duration of time spent with gaze directed towards the stimulus location, starting from the moment the owner's feet stopped moving on the X and continuing for ten seconds after.
Body angle	The angle of the dog's body in relation to the fence separating the stimulus presentation area, where 0° indicates that the spine was aligned directly towards the stimulus, and 180° indicates that the dog's body was facing directly away from the stimulus.
Gaze bouts	The number of times a dog looked in the direction of the stimulus presentation area during the ten seconds after the owner's feet stopped moving on the marker.
Average tail position	The average angle of the base of the dog's tail in relation to its spine during each 10 second interval. Coded on a Likert scale from 1-5: 1 = base of the tail is at or near 90-degree angle from the spine 2 = base of the tail is at a roughly 135-degree angle from the base of the spine 3 = base of tail is roughly even with the spine 4 = base of the tail is below the spine 5 = tail is tucked, with the base close to the dog's body.
Average body position	The average weight distribution of the dog during each 10 second interval, coded on a Likert scale from 1-3: 1 = weight forward, primarily on forepaws. 2 = weight neutral, equally spread on all four paws. 3 = weight backwards, primarily distributed to hind end.
Average mouth position	The average position of the dog's mouth during each 10 second interval (categorical): Closed, Open, or Moving (as in barking)
Average ear position	The average position of the dog's ears during each 10 second interval, coded on a Likert scale from 1-5: 1 = low/pinned 2 = half-low 3 = neutral 4 = half-high 5 = forward/high
Note. The behaviors above were coded as scaled variables.	

Table 4. Behaviors coded as present or absent at each distal marker

Behavior	**Description**
Trembling	Trembling/shaking movements of the body or head
Paw up	One front limb raised
Lip licking	Part of the tongue visible and moves along the upper lip
Yawning	Yawning
Panting	Breathing that is quick, audible, and visible
Whine	Whining
Low posture	Crouched posture
Pace	Pacing, walking at least one A→B →A route
Piloerection	Erection of fur along spine
Avoidance	Pressing against owner, dog positions itself so that the owner is between the stimulus and the dog, having first oriented to the stimulus
Grooming	Grooming
Freeze	Holding one position without any visible movement of the body while oriented towards the stimulus
Shake off	Oscillate vigorously the head and body on its longitudinal axis
Jump to Owner	Dog jumps towards something
Bowing	Lowering the forequarters while the hindquarters remain stationary
Paw at	Paw raised up and out in loose gesture
Wags tail	Repetitive wagging movements of the tail, to include only loose, gentle movements.
Exploration	Motor activity directed towards physical aspects of the environment (not the owner)
Orient to environment	Dog is still, but head is directed towards any object in the environment
Bark	Barking
Head tilt	Head tilt

Note. The behaviors above were coded as present or absent at each distal marker. In order to collapse across the distances, proportions of occurrences (out of 3 markers) were calculated.

Table 5. Behavior recorded while owners stood on the proximal marker

<table>
<tr><th>Category</th><th>Stimulus</th><th>Behavior code</th><th>Definition</th></tr>
<tr><td rowspan="2">Other</td><td rowspan="3">All stimuli</td><td>Environment</td><td>Duration spent sniffing the floor or objects in the environment that were not the stimulus.</td></tr>
<tr><td>Experimenter interaction</td><td>Duration spent sniffing an experimenter, or within 2ft of the experimenter with their gaze directed towards the experimenter</td></tr>
<tr><td>Owner</td><td>Owner interaction</td><td>Duration spent sniffing their owner, or within 2ft of the owner with their gaze directed towards the owner</td></tr>
<tr><td rowspan="4">Social</td><td rowspan="2">Model dog</td><td>Face sniffing</td><td>Duration spent with nose in contact with the model dog’s face, including the ears, muzzle, or any location anterior of the collar worn by the model dog.</td></tr>
<tr><td>Urogenital sniffing</td><td>Duration spent with nose in contact with the area directly beneath the tail or on the undercarriage of the model dog, immediately anterior to the hind legs</td></tr>
<tr><td rowspan="2">Novel object</td><td>Anterior</td><td>Duration spent with nose in contact with the narrow end of the novel object corresponding to where the model dog’s face was presented.</td></tr>
<tr><td>Posterior</td><td>Duration spent with nose in contact with the narrow end of the novel object corresponding to where the model’s hind end was presented.</td></tr>
<tr><td rowspan="2">Non-social</td><td rowspan="2">Model dog & Novel object</td><td>Proximal side</td><td>Duration spent with nose in contact with the side of the stimulus closest to the camera</td></tr>
<tr><td>Distal side</td><td>Duration spent with nose in contact with the side of the stimulus farthest from the camera</td></tr>
</table>

Figure 2. Average gaze duration at a distance by stimulus condition. A repeated measures ANOVA showed that average gaze duration varied significantly by stimulus condition ($F(3, 69) = 18.16, p < 0.0001$). Post-hoc pair-wise comparisons revealed that there was no significant difference between the control and novel object conditions, or between the model dog and live dog conditions. Dogs stared more in model dog and live dog conditions compared to the control and novel object conditions.

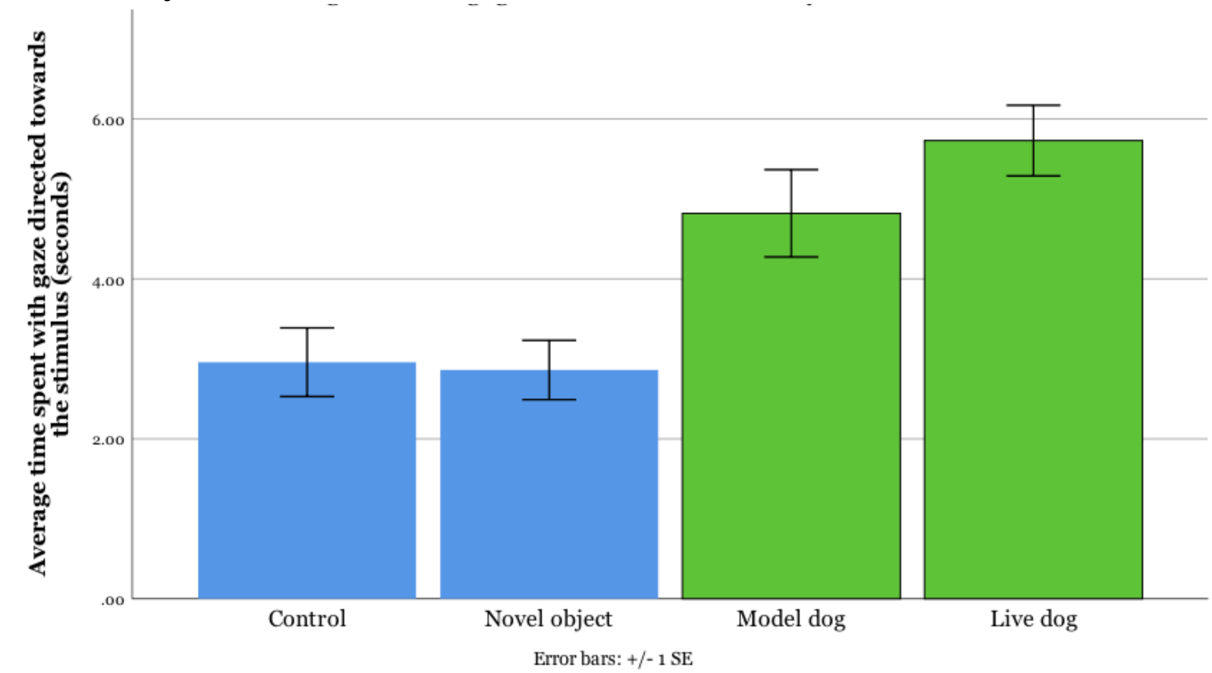


Table 6. Correlation matrix of dogs' average gaze duration by condition.

Stimulus condition		Control	Novel object	Model dog	Live dog
Control	Correlation Sig. (2-tail) N		**.523*** **.006** 25	.486 .012 25	.398 .049 24
Novel object	Correlation Sig. (2-tail) N			.287 .156 25	.307 .135 24
Model dog	Correlation Sig. (2-tail) N				**.729**** **<.0001** 24

Figure 3. Average time subjects spent sniffing a model dog compared to a novel object. There was no significant difference in average amount of time dogs spent sniffing between the two conditions (t (24) = 0.265, p = 0.793)

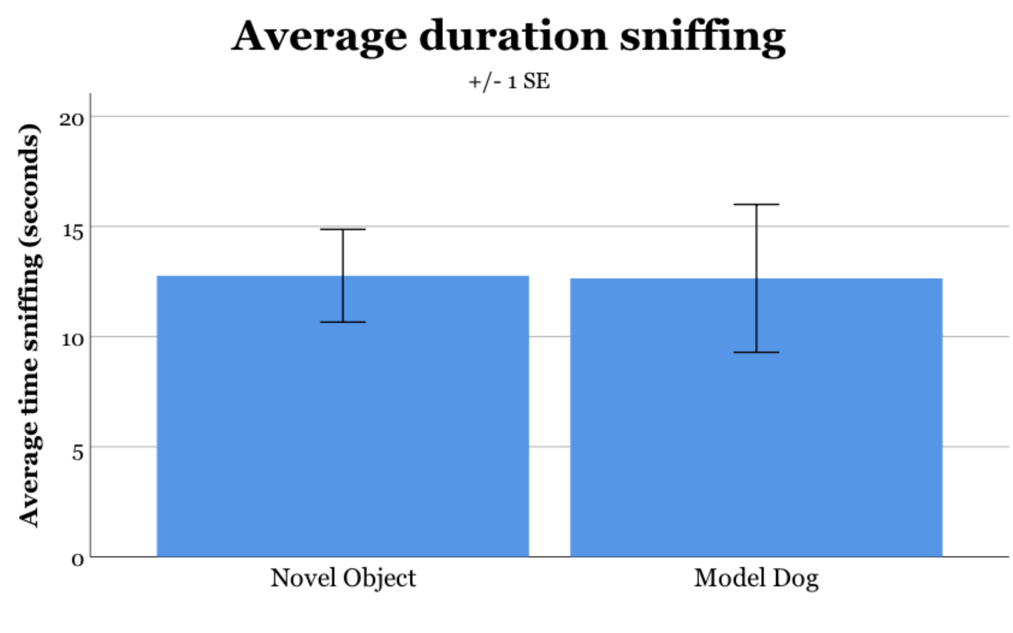


Figure 4. Proportion of total time sniffing stimuli that was directed toward socially relevant areas. We considered the face or urogenital region of the model dog socially relevant, as well as the corresponding locations on a novel object. Dogs spent significantly more of their total sniffing time on socially relevant areas of the model dog compared to the equivalent locations on the novel object. ($t(17) = 2.73$, $p = 0.014$).

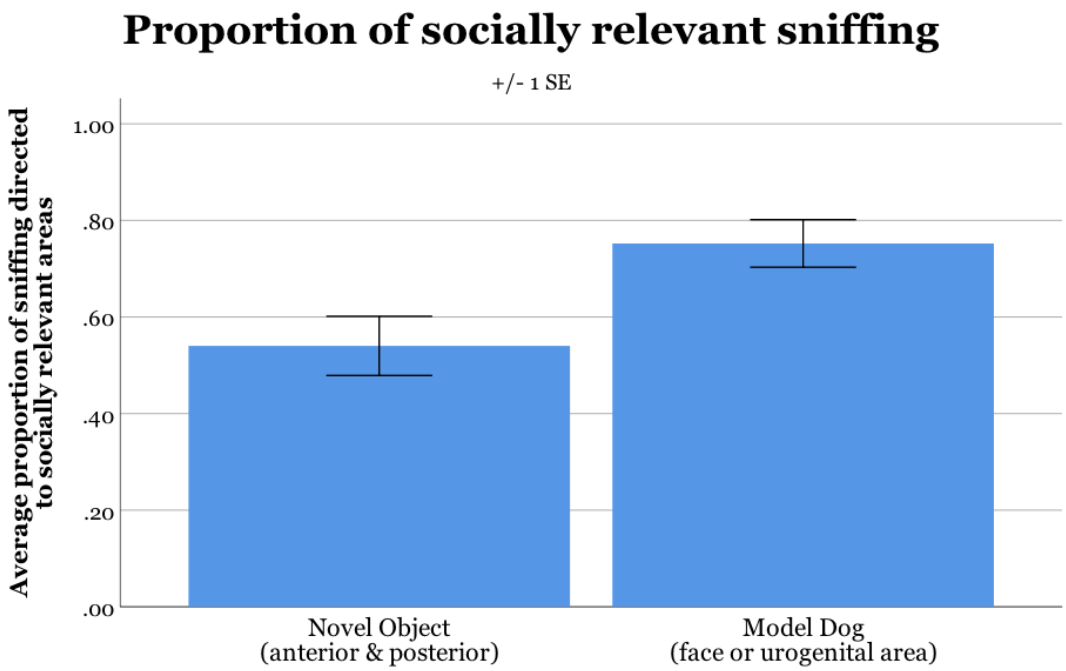


Figure 5. Hierarchical clustering of behaviors of interest. We used hierarchical clustering to group behaviors using the ICLUST package in RStudio. Orange indicates behaviors that were eliminated from final model selection analyses, while purple indicates behaviors that were clustered into a factor for further analysis. Factor labels are indicated on the left-hand side of the figure. Solid lines indicate a positive correlation between behaviors, while dashed lines indicate a negative correlation.

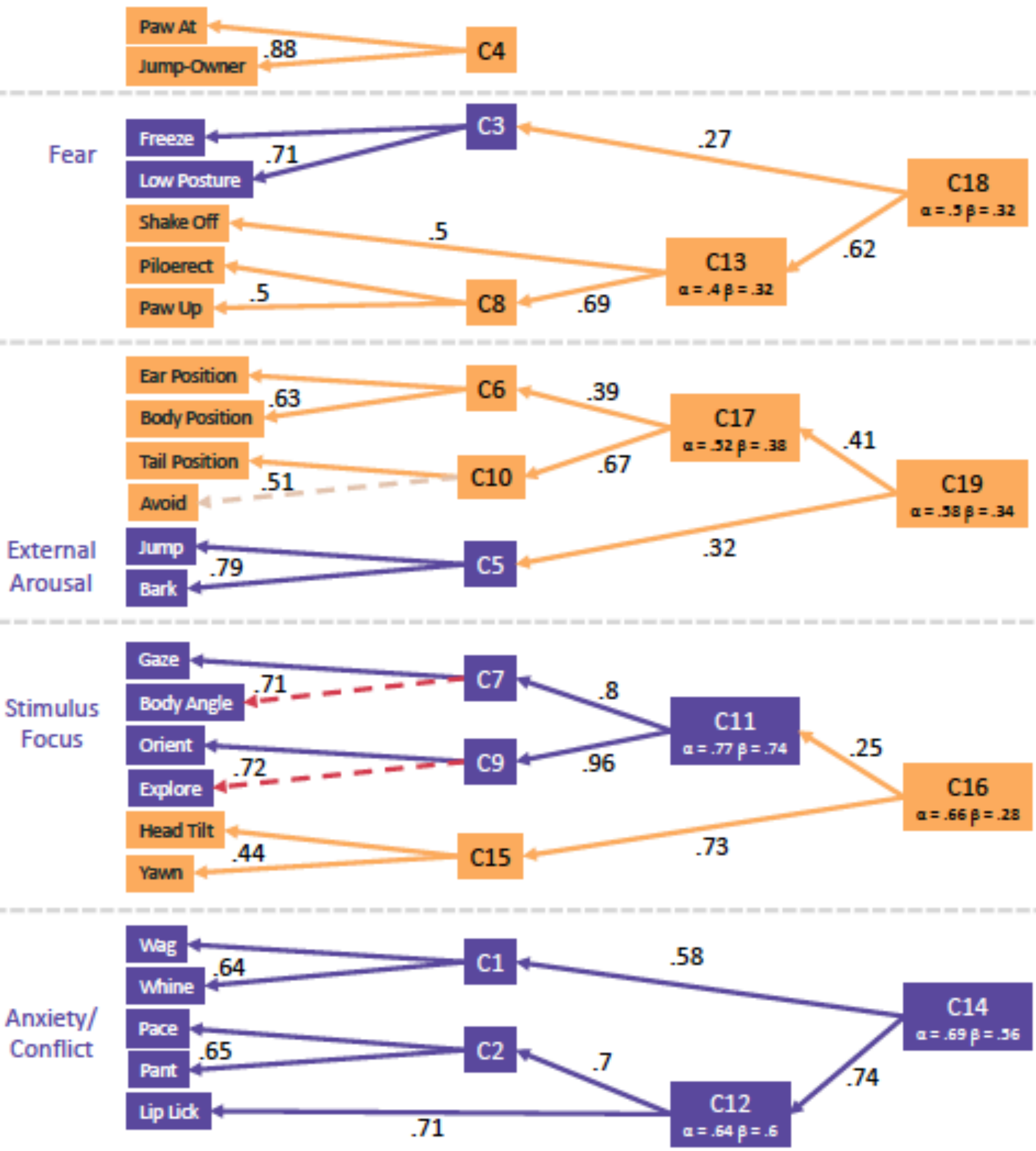


Figure 6. Correlogram of dogs' scores on each behavior cluster by condition. N = Novel object, M = Model Dog, L = Live Dog, C = Control. The correlations between dogs scores on each cluster in each condition, grouped by the strongest correlations. Blank squares are n.s. In each cluster, behavior was positively correlated across all conditions, meaning that dogs that were likely to perform behaviors in that cluster in one condition were more likely to perform them in another condition. In each behavior cluster, the behavior in the model dog condition showed the weakest correlation with behavior in the novel object condition.

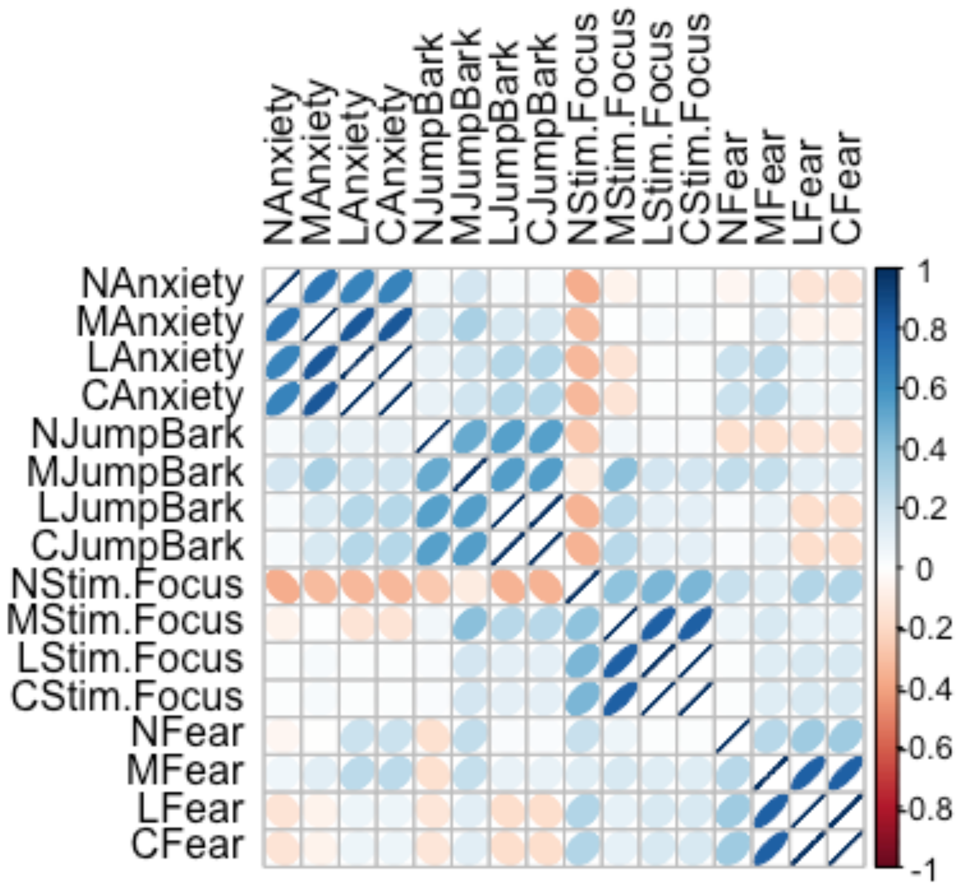


Table 7. AIC comparison of candidate models to predict Stimulus Focus scores.

Stimulus focus Candidate models	K	AICc	Delta AICc	AICcWt	AICcCum.Wt	Res.LL
Stimulus Condition	**6**	**200.57**	**0.00**	**0.65**	**0.65**	**-93.79**
AggrCBARQ	**10**	**202.04**	**1.47**	**0.31**	**0.97**	**-89.64**
DDAggrCBARQ	10	206.87	6.30	0.03	0.99	-92.06
NonSocialFearCBARQ	10	212.36	11.79	0.00	1.00	-94.81
Null Model	3	212.97	12.40	0.00	1.00	-103.35
FearCBARQ	10	213.39	12.82	0.00	1.00	-95.32
Excitability	10	214.34	13.77	0.00	1.00	-95.80
TrainabilityCBARQ	10	214.93	14.36	0.00	1.00	-96.09
CompositeArousal	10	215.84	15.26	0.00	1.00	-96.54
DDFearCBARQ	10	216.03	15.45	0.00	1.00	-96.64
SRBBARQ	10	216.61	16.04	0.00	1.00	-96.93
MiscCBARQ	10	216.99	16.42	0.00	1.00	-97.12
AttachCBARQ	10	217.48	16.91	0.00	1.00	-97.37

Note. All models contained Dog ID as a random effect. The Null Model contained only this random effect. The Stimulus Condition model also included contrasts to compare dogs' behavior based on the experimental condition. Models that end with "CBARQ" included the stimulus condition as well as the interaction between the conditions and dogs' scores on that particular C-BARQ scale.

Figure 7. Stimulus focus by stimulus condition. Dogs showed a progressive increase from the novel object to model dog to live dog conditions, meaning that they stared longer, oriented their body more directly, and explored less across these conditions.

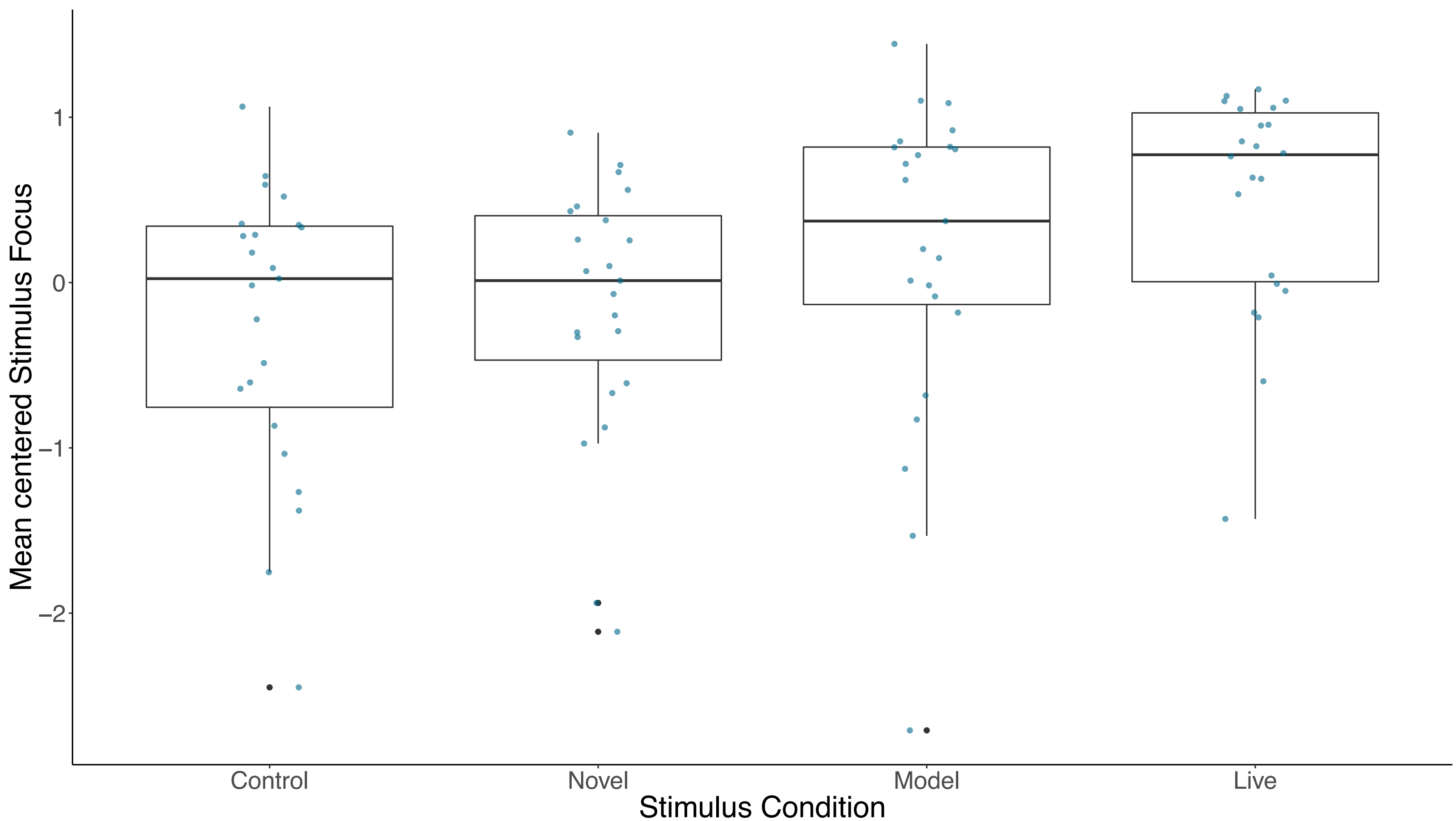


Figure 8. Stimulus focus predicted by stimulus condition and C-BARQ Aggression score. Dogs that scored low on aggression showed minimal differences in stimulus focus across conditions. Dogs with high C-BARQ Aggression scores oriented more toward the stimulus in the model dog and live dog conditions.

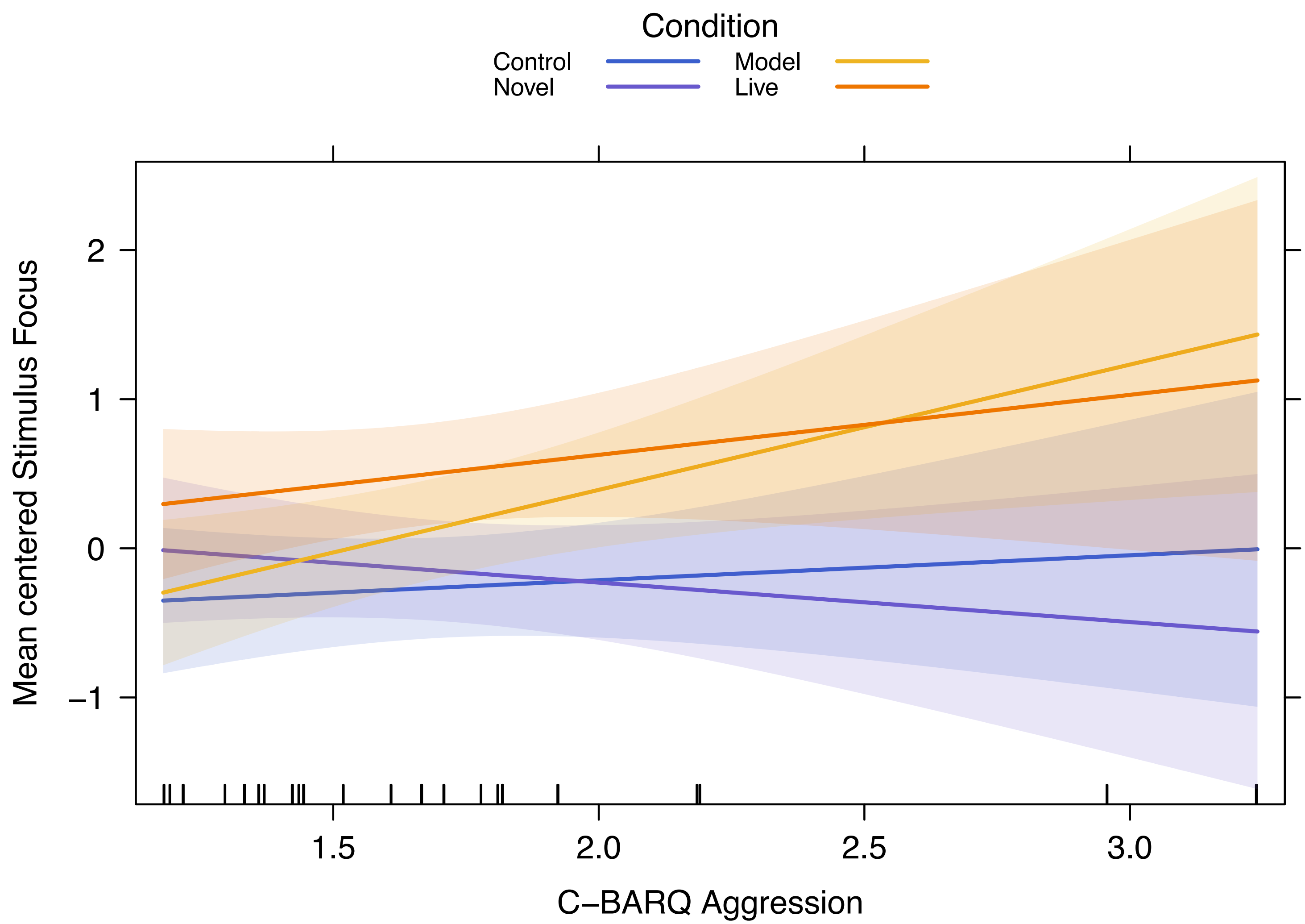


Table 8. AIC comparison of candidate models to predict External Arousal scores.

External Arousal Candidate models	K	AICc	Delta AICc	AICcWt	AICc Cum.Wt	Res.LL
Composite Arousal	**9**	**135.80**	**0.00**	**0.54**	**0.54**	**-57.79**
SRBCBARQ	9	138.15	2.35	0.17	0.71	-58.97
DDAggrCBARQ	9	138.97	3.16	0.11	0.82	-59.37
Stimulus Condition	5	140.21	4.40	0.06	0.88	-64.75
Aggression	9	140.70	4.89	0.05	0.93	-60.24
Excitability	9	142.09	6.29	0.02	0.95	-60.93
AttachCBARQ	9	142.54	6.73	0.02	0.97	-61.16
Null Model	2	143.37	7.57	0.01	0.98	-69.62
MiscCBARQ	9	144.08	8.27	0.01	0.99	-61.93
Trainability	9	146.27	10.47	0.00	0.99	-63.03
DDFearCBARQ	9	146.59	10.79	0.00	1.00	-63.19
FearCBARQ	9	146.80	10.99	0.00	1.00	-63.29
NonSocialFear	9	148.14	12.34	0.00	1.00	-63.96

Note. All models contained Dog ID as a random effect. The Null Model contained only this random effect. The Stimulus Condition model also included contrasts to compare dogs' behavior based on the experimental condition. Models that end with "CBARQ" included the stimulus condition as well as the interaction between the conditions and dogs' scores on that particular C-BARQ scale. Model(s) selected are in bold.

Figure 9. External Arousal by stimulus condition and Composite Arousal score. Dogs were more likely to jump and bark when a stimulus was present compared to the control condition. The model dog condition elicited the most extreme behavior, followed by the live dog and then the novel object conditions. Dogs that scored higher on Composite Arousal were more likely to jump and bark overall, and showed less discrimination in their behavior across conditions.

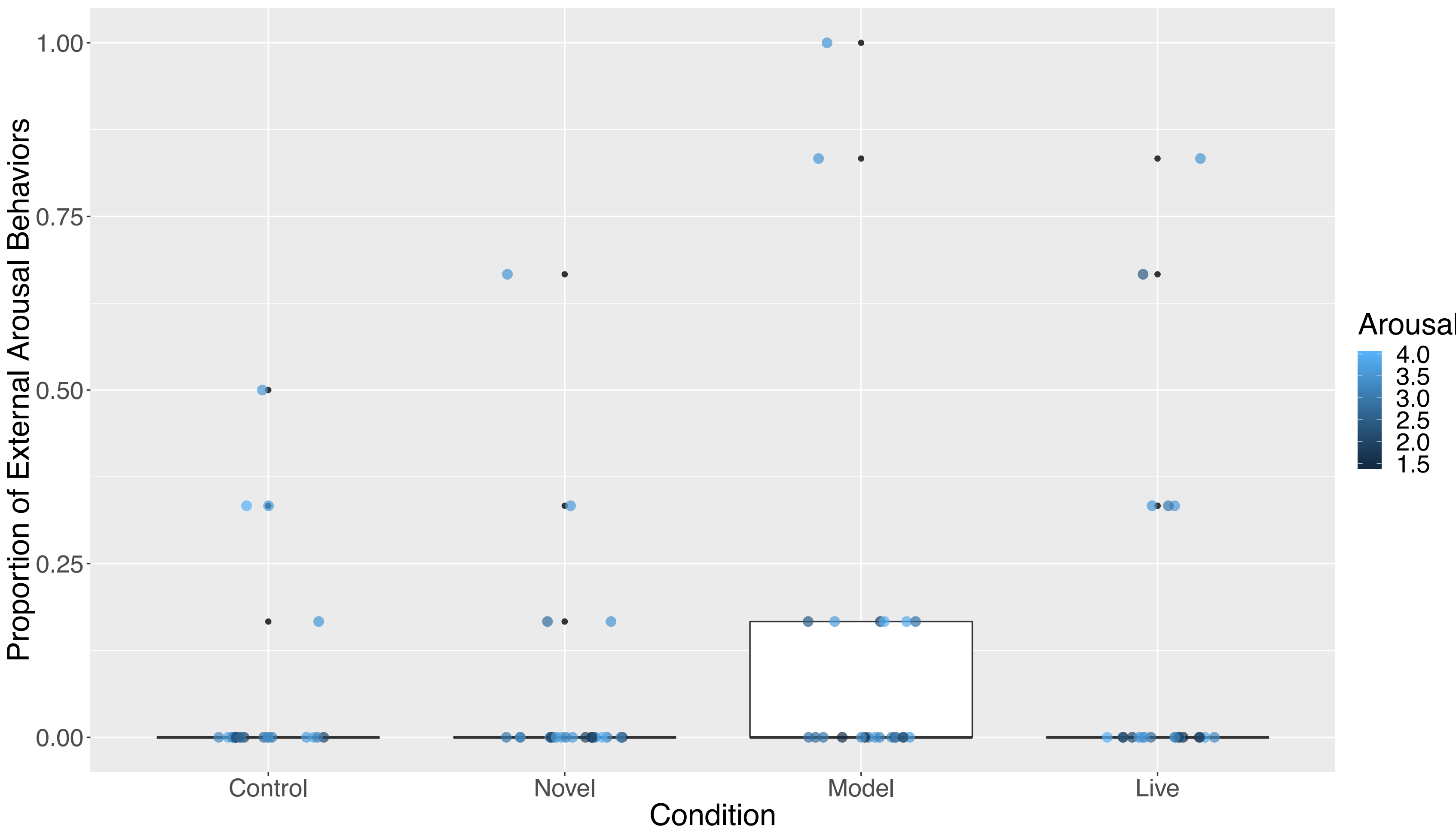


Table 9. AIC comparison of candidate models to predict Anxiety scores.

Anxiety Candidate models	K	AICc	Delta AICc	AICcWt	AICcCum.Wt	Res.LL
Stimulus Condition	**5**	**334.61**	**0.00**	**0.71**	**0.71**	**-161.95**
MiscCBARQ	9	338.22	3.62	0.12	0.83	-159.00
Fear	9	339.82	5.21	0.05	0.88	-159.80
NonSocial Fear	9	340.84	6.24	0.03	0.91	-160.31
CompositeArousal	9	342.35	7.74	0.01	0.93	-161.06
DDAggr	9	342.37	7.76	0.01	0.94	-161.07
Excitability	9	342.71	8.11	0.01	0.95	-161.25
Aggression	9	342.83	8.22	0.01	0.96	-161.30
Attachment	9	342.95	8.35	0.01	0.97	-161.37
DDFear	9	343.25	8.64	0.01	0.98	-161.51
SRB	9	343.68	9.08	0.01	0.99	-161.73
Trainability	9	343.78	9.17	0.01	1.00	-161.78
Null Model	2	348.12	13.52	0.00	1.00	-171.99

Note. All models contained Dog ID as a random effect. The Null Model contained only this random effect. The Stimulus Condition model also included contrasts to compare dogs' behavior based on the experimental condition. Models that end with "CBARQ" included the stimulus condition as well as the interaction between the conditions and dogs' scores on that particular C-BARQ scale. Model(s) selected are in bold.

Figure 10. Anxiety behaviors by stimulus condition. Dogs' tendency to wag, whine, pant, pace, and liplick was best predicted by the experimental condition. There was a progressive increase in the probability that these behaviors were exhibited from the control condition to the novel object to the model dog to the live dog condition.

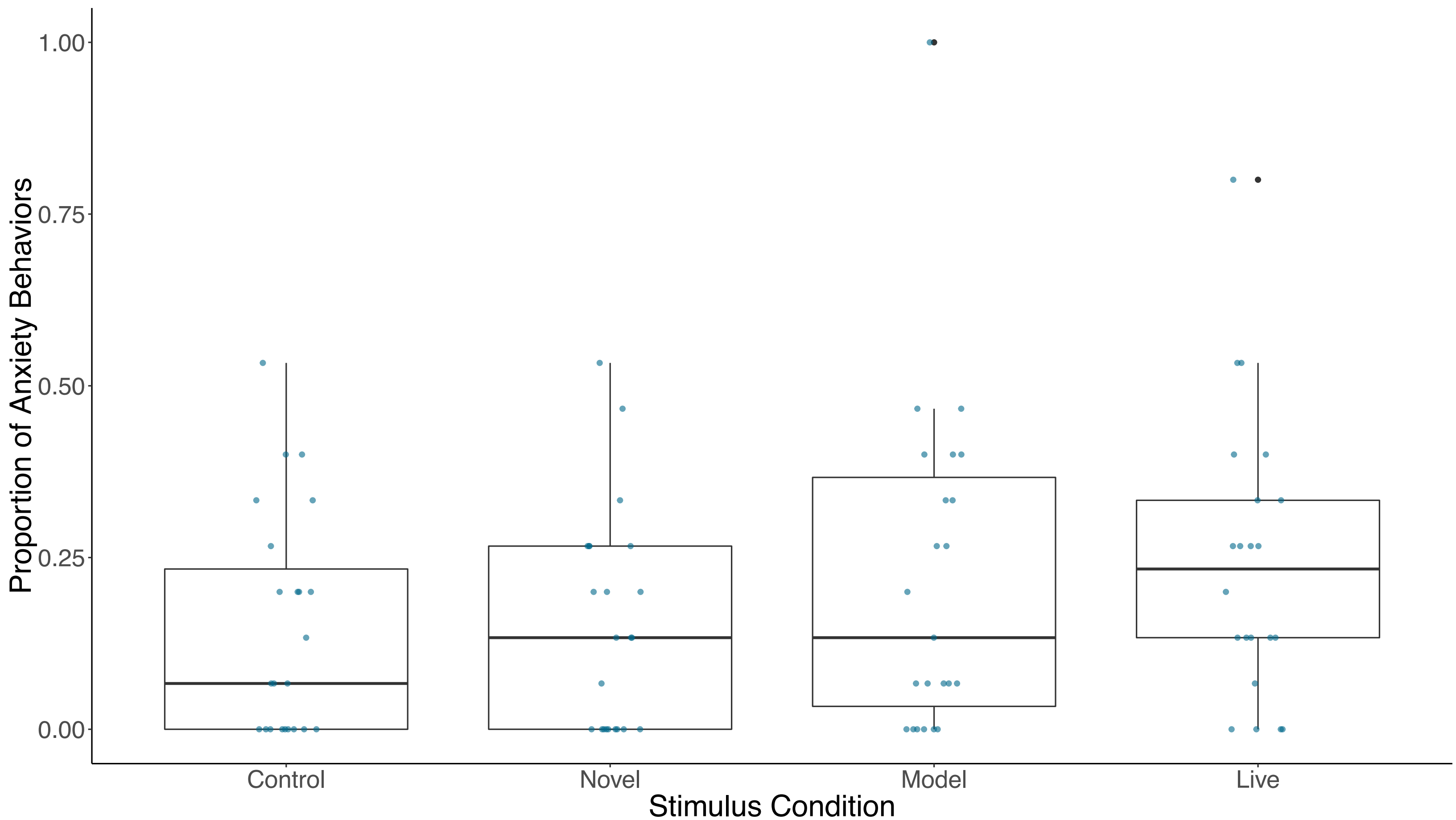


Table 10. AIC comparison of candidate models to predict Fear behaviors.

Fear Candidate models	K	AICc	Delta AICc	AICcWt	AICcCum.Wt	Res.LL
Null Model	**2**	**73.30**	**0.00**	**0.65**	**0.65**	**-34.58**
DDAggrCBARQ	9	75.92	2.62	0.18	0.83	-27.85
MiscCBARQ	9	78.63	5.34	0.05	0.87	-29.21
AggrCBARQ	9	79.18	5.88	0.03	0.91	-29.48
Stimulus Condition	5	79.55	6.25	0.03	0.94	-34.42
CompositeArousal	9	79.73	6.43	0.03	0.96	-29.75
TrainabilityCBARQ	9	80.66	7.36	0.02	0.98	-30.22
Fear	9	82.51	9.21	0.01	0.99	-31.14
Excitability	9	83.18	9.88	0.00	0.99	-31.48
Attachment	9	83.35	10.06	0.00	1.00	-31.57
NonSocial Fear	9	84.09	10.79	0.00	1.00	-31.93
SRB	9	85.59	12.29	0.00	1.00	-32.68
DDFear	9	88.11	14.81	0.00	1.00	-33.94

Note. All models contained Dog ID as a random effect. The Null Model contained only this random effect. The Stimulus Condition model also included contrasts to compare dogs' behavior based on the experimental condition. Models that end with "CBARQ" included the stimulus condition as well as the interaction between the conditions and dogs' scores on that particular C-BARQ scale. Model(s) selected are in bold.

3.6 REFERENCES

Barnard, S., Siracusa, C., Reisner, I., Valsecchi, P., & Serpell, J. A. (2012). Validity of model devices used to assess canine temperament in behavioral tests. *Applied Animal Behaviour Science*, pp. 79–87. https://doi.org/10.1016/j.applanim.2012.02.017

Barnard, S., Wells, D. L., Milligan, A. D. S., Arnott, G., & Hepper, P. G. (2018). Personality traits affecting judgement bias task performance in dogs (Canis familiaris). *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-25224-y

Beerda, B., Schilder, M. B. H., Van Hooff, J. A. R. A. M., & De Vries, H. W. (1997). Manifestations of chronic and acute stress in dogs. *Applied Animal Behaviour Science*, *52*(3–4), 307–319. https://doi.org/10.1016/S0168-1591(96)01131-8

Beerda, B., Schilder, M. B. H., Van Hooff, J. A. R. A. M., De Vries, H. W., & Mol, J. A. (1998). Behavioural, saliva cortisol and heart rate responses to different types of stimuli in dogs. *Applied Animal Behaviour Science*, *58*(3–4), 365–381. https://doi.org/10.1016/S0168-1591(97)00145-7

Bennett, S. L., Litster, A., Weng, H. Y., Walker, S. L., & Luescher, A. U. (2012). Investigating behavior assessment instruments to predict aggression in dogs. *Applied Animal Behaviour Science*, *141*(3–4), 139–148. https://doi.org/10.1016/j.applanim.2012.08.005

Borg, J. A. M. Van Der, Schilder, M. B. H., Vinke, C. M., & Vries, H. De. (2015). Dominance in Domestic Dogs : A Quantitative Analysis of Its Behavioural Measures, 1–18. https://doi.org/10.1371/journal.pone.0133978

Cannas, S., Frank, D., Minero, M., Aspesi, A., Benedetti, R., & Palestrini, C. (2014). Video analysis of dogs suffering from anxiety when left home alone and treated with clomipramine. *Journal of Veterinary Behavior: Clinical Applications and Research*, *9*(2), 50–57. https://doi.org/10.1016/j.jveb.2013.12.002

Christensen, E., Scarlett, J., Campagna, M., & Houpt, K. A. (2007). Aggressive behavior in adopted dogs that passed a temperament test. *Applied Animal Behaviour Science*, *106*(1–3), 85–95. https://doi.org/10.1016/j.applanim.2006.07.002

Deldalle, S., & Gaunet, F. (2014). Effects of 2 training methods on stress-related behaviors of the dog (Canis familiaris) and on the dog-owner relationship. *Journal of Veterinary Behavior: Clinical Applications and Research*, *9*(2), 58–65. https://doi.org/10.1016/j.jveb.2013.11.004

Diesel, G., Brodbelt, D., & Pfeiffer, D. U. (2008). Reliability of assessment of dogs' behavioural responses by staff working at a welfare charity in the UK. *Applied Animal Behaviour Science*, *115*(3–4), 171–181. https://doi.org/10.1016/j.applanim.2008.05.005

Döring, D., Roscher, A., Scheipl, F., Küchenhoff, H., & Erhard, M. H. (2009). Fear-related behaviour of dogs in veterinary practice. *Veterinary Journal*, *182*(1), 38–43. https://doi.org/10.1016/j.tvjl.2008.05.006

Hsu, Y., & 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*(9), 1293–1300. https://doi.org/10.2460/javma.2003.223.1293

Hsu, Y., & Sun, L. (2010). Factors associated with aggressive responses in pet dogs. *Applied Animal Behaviour Science*, *123*(3–4), 108–123. https://doi.org/10.1016/j.applanim.2010.01.013

MacLean, E. L., Gesquiere, L. R., Gruen, M. E., Sherman, B. L., Martin, W. L., & Carter, C. S. (2017). Endogenous oxytocin, vasopressin, and aggression in domestic dogs. *Frontiers in*

Psychology, *8*(SEP). https://doi.org/10.3389/fpsyg.2017.01613
Marder, A. R., Shabelansky, A., Patronek, G. J., Dowling-Guyer, S., & D'Arpino, S. S. (2013). Food-related aggression in shelter dogs: A comparison of behavior identified by a behavior evaluation in the shelter and owner reports after adoption. *Applied Animal Behaviour Science*, *148*(1–2), 150–156. https://doi.org/10.1016/j.applanim.2013.07.007
Mornement, K. M., Coleman, G. J., Toukhsati, S., & Bennett, P. C. (2010). A review of behavioral assessment protocols used by australian animal shelters to determine the adoption suitability of dogs. *Journal of Applied Animal Welfare Science*, *13*(4), 314–329. https://doi.org/10.1080/10888705.2010.483856
Mornement, K. M., Coleman, G. J., Toukhsati, S., & Bennett, P. C. (2014). Development of the behavioural assessment for re-homing K9's (B.A.R.K.) protocol. *Applied Animal Behaviour Science*, *151*, 75–83. https://doi.org/10.1016/j.applanim.2013.11.008
Mornement, K. M., Coleman, G. J., Toukhsati, S. R., & Bennett, P. C. (2015). Evaluation of the predictive validity of the Behavioural Assessment for Re-homing K9's (B.A.R.K.) protocol and owner satisfaction with adopted dogs. *Applied Animal Behaviour Science*, *167*, 35–42. https://doi.org/10.1016/j.applanim.2015.03.013
Neitz, J., Geist, T., & Jacobs, G. H. (1989). Color Vision In The Dog. *Visual Neuroscience*, *3*(2), 119–125. https://doi.org/10.1017/S0952523800004430
Ogata, N., Kikusui, T., Takeuchi, Y., & Mori, Y. (2006). Objective measurement of fear-associated learning in dogs. *Journal of Veterinary Behavior: Clinical Applications and Research*, *1*(2), 55–61. https://doi.org/10.1016/j.jveb.2006.06.002
Ottenheimer Carrier, L., Cyr, A., Anderson, R. E., & Walsh, C. J. (2013). Exploring the dog park: Relationships between social behaviours, personality and cortisol in companion dogs. *Applied Animal Behaviour Science*, *146*(1–4), 96–106. https://doi.org/10.1016/j.applanim.2013.04.002
Shabelansky, A., Dowling-Guyer, S., Quist, H., D'Arpino, S. S., & McCobb, E. (2015). Consistency of shelter dogs' behavior toward a fake versus real stimulus dog during a behavior evaluation. *Applied Animal Behaviour Science*, *163*, 158–166. https://doi.org/10.1016/j.applanim.2014.12.001
Sommerville, R., O'Connor, E. A., & Asher, L. (2017). Why do dogs play? Function and welfare implications of play in the domestic dog. *Applied Animal Behaviour Science*, *197*(March), 1–8. https://doi.org/10.1016/j.applanim.2017.09.007
Valsecchi, P., Barnard, S., Stefanini, C., & Normando, S. (2011). Temperament test for re-homed dogs validated through direct behavioral observation in shelter and home environment. *Journal of Veterinary Behavior: Clinical Applications and Research*, *6*(3), 161–177. https://doi.org/10.1016/j.jveb.2011.01.002

Supplementary

SI Table 1. Pictures that were used more than once in the body language questionnaire presented to dog owners.

Scale (Item #)	Scale pt represented	Second pairing (Item #)	Scale pt represented
Shy - Bold (3)	Bold (vs. neutral)	Low-High (9)	Neutral (vs. low)
Shy - Bold (4)	Neutral (vs. neutral)	Shy-Bold (7)	Neutral (vs. shy)
Shy - Bold (4)	Neutral (vs. neutral)	Low-High (4)	Low arousal (vs. high)
Friendly - Aggressive (8)	Neutral (vs. friendly)	Friendly - Aggressive (10)	Neutral (vs. neutral)
Friendly - Aggressive (8)	Friendly (vs. neutral)	Low-High (4)	High arousal (vs. neutral)

Chapter 4. COGNITIVE BIAS AND INTRASPECIFIC SOCIAL BEHAVIOR IN THE DOMESTIC DOG

Carolyn J. Loyer[1], James C. Ha[1], and Robin L. Foster[1,2]
[1]University of Washington, Seattle, Washington, USA
[2]University of Puget Sound, Tacoma, Washington, USA

Abstract

Research on individual differences in human behavior has repeatedly found that cognitive and attentional biases are related to mood and behavior disorders. In nonhuman animals, both chronic and acute environmental differences can impact cognitive biases, specifically judgement bias (JB). JB refers to how an individual responds to an ambiguous stimulus: as if it predicts a reward (a positive JB) or a punishment or non-reward (a negative JB). In domestic dogs (*Canis familiaris*), a negative JB was associated with a tendency to display more frequent or severe signs of distress when isolated from humans. In the present study, we examined whether owner reports of dogs' behavior via the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) predicted dogs' JB in an experimental context using a spatial discrimination task. Thirty-nine dogs were selected based on two C-BARQ criteria: high/low scores on the Fear and Anxiety scale (Fear), and high/low scores on questions relating to dog-directed Aggression (DDA). Higher Fear scores predicted a smaller decrease in latency to approach stimuli that were placed closer to a positive conditioned stimulus (CS+) location. Conversely, higher DDA scores predicted a more positive JB overall, and a sharper increase in speed to approach stimuli as they were closer to the CS+. A similar effect was found for Excitability. Although Excitability and Fear scores were moderately positively correlated, they had opposing effects on dogs' approach behavior to stimuli. These findings have implications for the treatment of canine behavior problems, particularly DDA. Our results suggest that treatment of DDA may benefit from

incorporating techniques that reduce arousal and discourage quick approach behavior toward conspecifics.

4.1 INTRODUCTION

Canine aggression can occur in a variety of contexts. Borchelt (1983) characterized eight categories of aggression in domestic dogs, while some clinical animal behaviorists distinguish between thirteen aggression types (Flint, Coe, Serpell, Pearl, & Niel, 2017). Further, there is considerable variability in terms of the predictability and severity of canine aggression (Netto & Planta, 1997; Sherman, Reisner, Taliaferro, & Houpt, 1996). Previous researchers have sought to understand the factors that predict both the risk of a dog exhibiting aggressive behavior, as well as the severity of aggression should it occur. A variety of factors have been implicated, such as sex, neuter status, physical condition, and rearing environment (Appleby, Bradshaw, & Casey, 2002; Borchelt, 1983; Casey, Loftus, Bolster, Richards, & Blackwell, 2014; Line & Voith, 1986; Sherman et al., 1996). Even individual differences in non-social behavior may be associated with aggression. For example, Flint et al. (2017) found a U-shaped relationship between non-social fear and the risk of a dog displaying severe stranger-directed aggression: Dogs with a moderate level of non-social fear were at lower risk of displaying severe aggression towards strangers. Conversely, dogs that displayed either no or severe levels of non-social fear were at a higher risk for displaying severe aggression toward strangers.

Recently, there has been increasing interest in the cognitive underpinnings of canine behavior problems, including aggression. In humans, cognitive biases (the tendency to focus on, notice, or interpret stimuli from the external environment in a biased manner) impact social behavior (Purkiss, Perrewé, Gillespie, Mayes, & Ferris, 2006), and individuals with higher levels of anxiety pay more attention to threatening stimuli than their non-anxious counterparts

(Goodwin, Yiend, & Hirsch, 2017). If individual differences in cognitive biases affect the way humans respond to their environment, we may expect that the same is true for non-human animals.

Judgement bias is one type of cognitive bias, and refers to an individual's response to an ambiguous stimulus: do they treat it as if it predicts a reward or positive outcome (a positive judgement bias) or as if it predicts an aversive or negative outcome (a negative judgement bias) (O. Burman et al., 2011). Researchers have investigated the situations that predict positive and negative judgement biases in dogs using discrimination tasks in which dogs are conditioned to expect a reward paired with one stimulus (the CS+), and non-reward or aversive with another (the CS-) (e.g. Barnard, Wells, Hepper, & Milligan, 2017; O. Burman et al., 2011; Mendl et al., 2010a). Once subjects are consistently approaching or responding to the CS+ and avoiding or delaying approach to the CS-, they are presented with ambiguous stimuli that are intermediate between the two conditioned stimuli. A positive judgement bias is demonstrated by a response that is more similar to the individual's speed of approach or response towards the CS+ than the CS-, and the opposite for a negative judgement bias. This can be done using auditory, visual, or spatial paradigms.

Using a spatial discrimination task, Mendl et al. (2010) found that a negative judgement bias correlated with higher levels of separation-related behaviors (SRB) in dogs in a shelter. A separate study found that treating companion dogs for SRB with fluoxetine and behavior modification induced a more positive judgement bias (Karagiannis, Burman, & Mills 2015). These studies provide evidence that there is a relationship between judgement bias and the tendency to show signs of distress in the absence of a familiar human. It is likely that judgement bias is related to other behavior differences as well.

Barnard, Wells, Milligan, Arnott, & Hepper (2018) investigated the relationship between JB and a variety of individual differences. They found that dogs that are more excitable, and those that tend to show fear in non-social situations tended to have a more positive judgement bias, whereas dogs with higher degrees of SRB and those that showed more fear or aggression towards other dogs had a more negative judgement bias. One limitation proposed by the study's authors is that the behavior range of their sample may have been restricted, especially at the extremes of aggression and fear.

The restriction of range in terms of subject behavior is a common limitation in behavior research. Individual differences in behavior can influence sampling when wild subjects are selected by trapping (Biro & Dingemanse, 2016, but see also Michelangeli, Wong, & Chapple, 2016). In addition, when participation is voluntary, personality influences both willingness to participate as well as success on cognitive tasks (Morton, Lee, & Buchanan-Smith, 2013). The problem of a limited behavioral range in the sample is particularly understandable in research using companion dogs as subjects. Companion dog owners may be reluctant to bring their behaviorally challenged pet to participate. Additionally, conducting research with dogs with a history of fear and/or aggression raises welfare concerns, as well as a risk of injury to the researchers. However, taking steps to overcome this range restriction is imperative in order to better understand the causes of behavior problems in companion dogs, and to improve our methods for treating them.

In the present study, we examined the relationship between JB and two common canine behavior problems, fear and dog-directed aggression (DDA), in companion dogs that represented a wide range of the behavioral spectrum on both dimensions. Secondarily, we analyzed whether relationships between JB and other individual differences found in previous studies were replicated in our sample (e.g. excitability, SRB, etc.). We hypothesized that, if the location-based judgement bias task is a valid metric of judgement bias, higher levels of fear would be related to a more negative judgement bias. Fear may be broadly defined as a tendency to approach uncertain situations as if they may be aversive or unsafe, so we would expect that dogs whose owners report they are more fearful in their daily lives would also treat ambiguous stimuli in an experimental setting with a more negative JB. Conversely, the relationship between DDA and JB may be more complicated.

Previous research has demonstrated that in dogs, fear-based aggression is common, but also that not all aggression is motivated by fear (Cannas et al., 2018; Fatjo, Amat, Mariotti, de la Torre, & Manteca, 2007). In wild animals, competition for resources such as territory, food, and mates may lead to agonistic interactions. Initiators of these conflicts may do so in anticipation of a positive outcome (e.g., obtaining resources), or demonstrating a positive JB. Thus, we were unsure whether DDA would be related to judgement bias.

4.2 METHODS

4.2.1 Ethical statement

All research protocols were approved by the University of Washington IACUC under protocol 4413-02-201800018. Dog owners gave informed consent prior to their dog's participation, and were present throughout the experiments.

Subjects

Dog owners were recruited via social media, word of mouth, and local trainers' and behavior consultants' clients. Interested owners completed an online intake questionnaire, which included demographic information regarding their dog, screened for severe aggression towards humans, and screened for any physical limitations that may limit their dog's ability to participate in the study.

The questionnaire also included the Canine Behavioral Assessment and Research Questionnaire (C-BARQ) (Hsu & Serpell, 2003), and the Canine Body Language Questionnaire (CBLQ) (Loyer & Foster, in prep). For the purposes of this study, we will only examine C-BARQ data, as the C-BARQ has been previously validated while the CBLQ is still in development. We received 229 complete survey responses. Of these surveys, 206 dog owners were interested in further participation. We calculated scores for general fear using the average score owners reported on the Fear and Anxiety portion of the C-BARQ, and scores for dog-directed aggression (DDA) using the average of a subset of questions from the Aggression portion of the C-BARQ. Specifically, we included only questions that dealt with a dog's tendency to display aggression towards other dogs. Questions that pertained to aggression directed towards humans were not used in this score. Scores were calculated by a research assistant who then anonymized the questionnaire responses for participant selection.

CL sorted dogs into five categories based on their DDA, and Fear. Dogs with a DDA score ≥ 2.5 were classified as "aggressive", while dogs with scores ≤ 2.0 were classified as "sociable". Similarly with Fear, dogs with a score ≥ 2.5 on the C-BARQ Fear and Anxiety scale were classified as "fearful", and dogs scoring ≤ 2.0 were classified as "confident." This classification system resulted in four categories: Sociable-Confident dogs, Sociable-Fearful dogs, Aggressive-Confident dogs, and Aggressive-Fearful dogs. Dogs receiving intermediate scores on either scale did not fit into one of the four categories and were excluded from in-person participation. Using the remaining four categories, we matched dogs as closely as possible by

American Kennel Club breed group, by age, and by size to ensure roughly balanced groups across the four categories.

4.2.2 Procedure

All sessions took place at a dog training facility in Seattle, WA (University Canine Learning Academy). The facility was 1180ft^2, with rubber mat flooring for traction. Prior to participation, owners were asked to fast their dogs for at least four hours to ensure that they were not satiated. Immediately prior to the session owners signed a consent form and each dog's leash was exchanged for a 5' nylon flat leash (.75" wide) with a padded handle. For smaller dogs, a lighter, narrower leash (.5") of the same length and design was used. Each session was divided into three portions: habituation, learning, and testing.

During the habituation phase, dogs were given at least 5 minutes to explore and acclimate to the space. During this time, the researcher acting as the dog's handler (CL was the handler for all dogs) gave owners verbal instructions and an opportunity to ask questions. She also ensured that each dog would willingly eat food from a metal bowl, and that the dog would walk on lead with her. Dogs who showed fear-related behavior in either situation were given additional time to acclimate to the bowl or to walking with the handler. Because we recruited half of our subjects based on owners reporting high degrees of fear-related behavior, protocols were in place to habituate any dog that showed signs of fear such as trembling, panting, refusing food, and avoidance of the researchers or the testing area. One dog did not habituate to the metal bowl, so a paper cup was used instead.

After the habituation period, we proceeded to the spatial discrimination learning phase. Dogs and their owners were escorted behind a barrier where two chairs were stationed, along with a metal water bowl. CL and the owner sat in the chairs while a research assistant placed the stimulus (a metal bowl with boiled chicken breast in it was placed at the CS+ location--the side of the room decided by a coin toss, and the same metal bowl, empty, was placed in the CS- location on the opposite side of the room). When the bowl was in place and the RA had returned to their position behind the fence, the RA verbally cued CL to bring the dog to the testing area. CL brought the dog from behind the barrier and walked it to an X marked on the floor in masking tape before saying “Okay!” and dropping the dog’s leash.

For at least the first four learning trials, owners walked out with the dog and handler to minimize stress and ensure that the dogs would participate. Once the dog was walking comfortably with the handler, owners were asked to sit in one of the chairs with their back toward the testing area so that they could not observe the trials. Between each trial owners were welcome to interact with their dog as normal. Dogs received two pieces of their regular dry kibble or a less enticing treat each time they returned to the station. Some dogs did not habituate to leaving their owner to walk with the handler. For those dogs, the owner accompanied the dog and the handler for each trial. They were coached to stop behind the dog and handler, and to maintain a neutral expression and ignore the dog. Each trial lasted roughly 30 seconds, and each dog received a minimum of 15 learning trials.

The order of stimulus presentation was randomized in advance, and identical for all dogs (see supplementary material for the complete trial order). A coin toss at the start of each session determined which side would be the CS+ for the first participant. Thereafter, the CS+ side alternated for each subsequent dog.

Each trial proceeded as follows:

1) Research assistant cues handler that the bowl is placed.

2) Handler and dog walk out from behind the barrier towards the starting point marked on the floor.
3) Handler says “Okay!” and drops the leash, while looking at the bowl. RA starts the stopwatch as soon as the dog is released.
4) Dog is given up to 20 seconds to approach the bowl. RA stops the stopwatch when the dog’s nose passes the threshold of the bowl and records the latency to approach.
5) When the 20 seconds has passed, or the dog has moved away from the bowl, handler says “Good job!” and calls the dog back to return to the station with the owner.
6) Handler gives the dog two pieces of their regular dry food, and waits to be cued by the RA to repeat for the next trial.

During the discrimination learning phase, the only conditions were the CS+ and the CS-. We followed Mendl et al.'s (2010) criteria for learning, such that we considered that a dog had learned the task when the preceding three CS+ trials had lower latencies than the fastest of the three preceding CS- trials; i.e., dogs were faster to approach the CS+ than the CS- on at least six consecutive trials.

The testing phase began once the dog had met the learning criteria, and the owner and dog returned from a brief (2-5 minutes) walk break. First we conducted four refresher trials to ensure that there was no regression in learning. If the dogs were still faster to approach the CS+ than the CS-, we continued to the testing trials. If not, additional learning trials were conducted until the dog met criteria again.

During the testing trials, we followed the order of presentation of ambiguous stimuli used in Mendl et al.’s (2010) study. Three probe stimulus locations, Near Positive (NP), Middle (M) and Near Negative (NN), were presented in three testing trials each, with four refresher trials of the previously learned CS+ and CS-stimulus locations between each probe trial. The order of presentation during the refresher and testing trials was the same across all dogs (Table 2). The final trial for each dog was a control trial that consisted of placing the bowl in the CS+ location without food in it to ensure that dogs were not basing their decision to approach on visual or odor cues. Thus, including the refresher trials, the testing phase consisted of 47 total trials: 4 refresher trials, 12 probe trials, 30 trials using either the CS+ or CS- locations, and 1 control trial.

4.2.3 Data Analysis

We used survival analysis models to examine the effects of stimulus location, trial order, and dog C-BARQ scores on probability and latency of approach. Survival analysis measures the probability of an event occurring (or the proportion of a sample for which an event has occurred) across time. For example, the proportion of a population that has show symptoms of an illness 2, 4 or 10 days after first exposure. In our analysis, the event of interest was approaching the stimulus (the dog bowl, presented at the five various locations), and we examined how the latency and probability of approach varied by location as well as with individual differences in the subjects' behavior. As a result of the training trials, prior to testing all dogs demonstrated a shorter latency (and higher probability) of approach toward the CS+ location compared to the CS- location. Survival analysis in this case allowed us to investigate how dogs' approach behavior toward the ambiguous stimuli changed relative to their behavior toward the conditioned stimuli to determine whether they were exhibiting a positive judgement bias (probability and speed of approach near that of their approach to the CS+) or a negative judgement bias (approach behavior closer to that of the CS-).

We compared Cox Proportional Hazards models of approach behavior using Akaike Information Criterion (AIC) with a correction for small sample size, as well as Bayesian

Information Criterion (BIC) to ensure that we were not overfitting with multiple comparisons. First we examined the effect of probe characteristics on dogs' tendency to approach the bowl. We included models with the main effect of probe position (NNeg, M, NPos), probe order (whether it was the first, second, or third time that probe location had been presented), and whether the trial preceding the probe was a CS+ or CS-. Previous research has demonstrated that judgement bias is affected by events immediately preceding the measurement task (Bateson, Desire, Gartside, & Wright, 2011a; Doyle, Fisher, Hinch, Boissy, & Lee, 2010; Sanger, Doyle, Hinch, & Lee, 2011) as well as long-term living conditions (Bethell & Koyama, 2015; Brydges, Leach, Nicol, Wright, & Bateson, 2011; Douglas, Bateson, Walsh, Bédué, & Edwards, 2012; Parker, Paul, Burman, Browne, & Mendl, 2014). Thus, we tested whether dogs' approach behavior was predicted by the type of trial that immediately preceded the probe.

Subsequently, we examined the effects of intra-subject differences (C-BARQ scores) on latency to approach (see Table 1 for a complete list of the C-BARQ scales and subscales, and their descriptions).

4.3 RESULTS

Forty-five owners brought their dogs to participate. Of these, six were excluded (2 due to dogs barking or lunging at experimenters that did not abate adequately with the habituation period, 1 due to extreme fear that presented after the first CS- training trial, 2 due to physical limitations that were more extreme than disclosed in the intake questionnaire, and 1 because the dog became satiated before testing trials were complete), resulting in a final sample of thirty-nine dogs. Nineteen (49%) of these dogs were female, of which all but one were spayed. The remaining 20 (51%) participants were all neutered males. The average age (± SD) of the dogs was 4.30 years (± 2.48 years), and they did not differ significantly by sex ($t(34) = 0.147$, $p =$

0.884) or behavior category ($F(3,3\ 5) = .0.156$, $p = .925$) (Figure 1). On average, dogs had been living with their owners for 3.18 years (± 2.20 years). Duration owned did not differ by behavior category ($F(3, 35) = 0.992$, $p = .408$).

4.3.1 Learning trials to reach criterion

The average number of learning trials to reach criterion was 20.05 $\pm$ 4.88, with a maximum of 30 trials. There were no differences across behavior categories in average number of learning trials to reach discrimination criterion ($F(1, 35) = .001 - .103$, $p = .750 - .971$). Number of learning trials to reach criterion did not correlate with C-BARQ Trainability score ($r = .014$, $p = .9304$, Figure 3).

4.3.2 Differences in approach behavior based on stimulus characteristics

In general, dogs showed a positive JB. Collapsing across the probe stimuli, latencies to approach the ambiguous locations were closer to that of the CS+ than the CS- ($z = 7.67$, $p = 1.7 \times 10^{-14}$). Dogs were less likely to approach the ambiguous probe stimulus locations than the CS+ location, and those that did approach did so more slowly the farther the stimulus was from the CS+ location ($z = 34.42$, $p < .001$).

Dogs appeared to learn that probe stimuli locations were not associated with food. The likelihood of approaching and speed of approach decreased progressively the second and third times probe stimuli were presented ($z = 10.43$, $p < .00000001$, Figure 4), and this change was non-linear ($z = 3.55$, $p = 3.8 \times 10^{-4}$), meaning that there was a larger increase in latency from the first to the second presentation of probe stimuli than between the second and third presentations. The type of trial preceding the probe trial also affected dogs' latency to approach. If the previous trial was a CS+ (i.e., food was in the bowl), dogs were slower and less likely to approach the subsequent probe stimulus ($z = 8.05$, $p = 7.8 \times 10^{-16}$).

4.3.3 Differences in approach behavior based on C-BARQ scores

C-BARQ Trainability interacted with the effect of the previous trial (i.e., whether the preceding trial was a CS+ or a CS-) to affect approach behavior towards ambiguous stimuli. Dogs whose owners reported that they were more trainable showed a smaller effect of the stimulus immediately preceding the probe trial ($z = 2.04$, $p = 4.2 \times 10^{-2}$). These dogs also tended to have smaller difference in latency change between the first and second compared to second and third presentations of the probe stimuli ($z = 2.39$, $p = .017$), though this difference was not significant after controlling for multiple comparisons.

Excitability and SRB affected approach behavior: dogs with higher Excitability and higher SRB C-BARQ scores showed a stronger increase in speed to approach a stimulus the closer its position was to the CS+ ($z = 5.73$, $p = 1 \times 10^{-8}$ and $z = 3.54$, $p = 4.1 \times 10^{-4}$, respectively, see Figures 5 & 6). C-BARQ SRB scores affected dogs' behavior towards the probe stimuli. Dogs with higher average SRB scores were less likely to approach probe stimuli locations, and those that did approach did so more slowly than dogs with lower SRB scores ($z = 3.07$, $p = 2.1 \times 10^{-3}$), demonstrating a more negative JB (Figure 6). A higher SRB score also predicted less similar latencies to approach between the near-positive and near-negative probe locations compared to dogs with lower SRB scores ($z = 4.09$, $p = 0.000043$).

4.3.4 Dog-Dog Aggression & Fear

Dogs with higher C-BARQ DDA scores showed a more positive JB. Higher DDA C-BARQ scores predicted a shorter latency and increased probability of approach to probe stimulus locations ($z = 3.74$, $p = .00018$), and a stronger decrease in latency as the stimulus was closer to the CS+ location ($z = 2.62$, $p = .0089$, Figure 7).

There was an opposite effect of C-BARQ Fear scores on approach behavior across stimulus locations. Dogs that were higher in fear had a smaller decrease in latency as the stimulus moved from CS- to CS+ positions compared to dogs with lower levels of fear ($z = 5.68$, $p = 1.3 \times 10^{-8}$, Figure 8). Although Fear and Excitability scores were moderately positively correlated ($r = 0.47$, $p = 0.0024$), they had opposing effects on approach behavior based on probe position. There was not a significant effect of Fear scores on dogs' overall tendency or speed of approach towards probe stimuli versus the trained stimuli ($z = .74$, $p = 0.46$), but a higher Fear score predicted a smaller change in approach behavior between the near-positive and near negative probe locations ($z = 2.69$, $p = 0.0072$).

4.4 Discussion

Dogs learned quickly to discriminate between two locations, one associated with food (CS+), and the other with absence of food (CS-). This pattern is consistent with other studies using spatial discrimination to study JB (Barnard et al., 2018; Mendl et al., 2010a). There were no differences in average number of trials to reach criterion across the previously selected behavior categories, and C-BARQ Trainability scores were not associated with the number of trials required for a dog to reach criterion. In general, dogs showed a positive JB, and this effect was exaggerated for dogs whose owners reported that they show more severe signs of aggression towards other dogs, i.e. a high DDA score on the C-BARQ was associated with a more positive judgement bias. The opposite was true for SRB: a higher SRB score was associated with a more negative JB overall, and decreased difference between the NPos and NNeg locations.

Dogs that were more fearful as assessed by the C-BARQ showed a greater increase in latency as the stimulus location moved toward the CS-, particularly when comparing their latency and probability of approaching the NPos versus NNeg probe stimuli. Thus, more fearful dogs did not increase their speed of approach as stimuli moved closer to the CS+ location as much as confident dogs. Although C-BARQ Excitability scores were correlated with Fear scores, they had opposing effects on behavior. Dogs that were scored as more excitable showed a stronger change in behavior based on stimulus location than less excitable dogs.

On average, dogs' latency to approach the ambiguous stimuli was nearer that of their approach to the CS+ than the CS-, suggesting a generally positive JB. This finding is consistent with previous studies in which the difference between the conditioned stimuli is the presence versus absence of a reward, as opposed to the presence of a reward versus an aversive (Mendl, Burman, Parker, & Paul, 2009). This general positive bias may also be an artefact of the experimental set up. We used a regular metal dog bowl for all locations, which we assume had a pre-existing positive association for most of our subjects. Additionally, dogs received a small treat of their regular dry food between each trial which may have had a positive effect on their emotional state. Previous research has demonstrated that food rewards induce a positive JB in sheep (Verbeek, Ferguson, Quinquet de Monjour, & Lee, 2014). Thus, there may have been some carryover effect. Our results may have differed if we had included a deterrent such as a white noise burst when dogs approached the CS-. However, the use of such an aversive stimulus would likely have caused greater attrition due to the subjects refusing to participate, particularly since roughly half of the dogs were selected based on high levels of fear. One additional explanation may be that in the process of domestication, dogs have been selected for neophilia (a tendency to be attracted to novelty), resulting in an increased tendency to approach novelty.

In contrast to Barnard et al.'s (2018) study, which found a correlation between DDA and a more negative JB, we found that dogs whose owners reported higher levels of dog-directed aggression had a more positive judgement bias than other subjects. This finding has important implications for the treatment of this behavior challenge. Currently, desensitization and counterconditioning have the most scientific support as treatment methods for dogs who are reactive towards other dogs (Overall, 2013). The underlying assumption of counterconditioning is that aggression and reactivity towards conspecifics stems from a negative association with, or affective state in the presence of, other dogs. By pairing the stimulus (another dog) with a different, positively-valenced stimulus (e.g., a food reward), trainers and behavior consultants seek to change the valence of a dog's association with other dogs, and change their behavior as a result. In order to be effective, the positive stimulus used in counterconditioning must be of high enough value to counteract the negative association with the other trigger. In practice, this often means using novel treats and toys that are highly desired. While this should result in a shift in emotional valence, the introduction of these stimuli may also increase the degree of arousal a dog experiences in that context.

In our study, dogs that tended to be highly excitable showed a stronger increase in speed towards potentially rewarding stimuli than less excitable dogs. If high levels of arousal impact a dog's decision-making and tendency to approach stimuli, and dog-directed aggression is not necessarily associated with anticipation of a negative outcome, then treatment of dog reactivity should emphasize decreasing arousal and discouraging the tendency to immediately approach other dogs.

Our results support previous findings that individual differences in behavior relate to judgement bias (Barnard et al., 2017; Karagiannis et al., 2015; Mendl et al., 2010a). Dogs with higher Fear scores showed a decreased effect of stimulus location (i.e., a decreased preference for the CS+ over the CS-). Conversely, dogs that were higher in Excitability demonstrated an *increased* change in behavior based on stimulus location. The opposing effects of these scores are of particular interest, given that Fear and Excitability were positively correlated. Revisiting the questionnaire used to calculate these scores provides one potential explanation. Excitability is described as "becoming highly excited at the slightest novelty" and includes "movement towards the source of novelty" (Hsu & Serpell, 2003), and the questions used in its calculation ask about dogs' responses in situations such as when guests arrive, or when owners return after a brief absence. Similarly, fear is outlined as a tendency to crouch or cringe, whimper, or (in more severe cases) attempt to flee the triggering stimulus, such as in unfamiliar situations or when someone attempts to pet them. Both Excitability and Fear are characterized by an overreaction to introduced stimuli, therefore they may not be independent of each other, physiologically or motivationally.

The distinction between Fear and Excitability in their effects may be better understood in the context of the Reinforcement Sensitivity Theory of Personality (RST; Corr, 2008). The RST posits that many individual differences in behavior stem from biologically based differences in sensitivity to rewards, and to threat or punishment, as well as differences in how conflict between the two is mediated. The RST identifies three main systems: – the Behavioral Activation System (BAS), the Fight/Flight/Freeze System (FFFS), Behavioral Inhibition System (BIS). The BAS's primary role is in detecting and responding to stimuli that are rewarding o that predict rewards or positive outcomes. The FFFS is essentially the opposite: It is responsible for actively avoiding punishing or threatening situations. The BIS, however, acts as the mediator between the BAS and the FFFS when a situation has the potential to produce *either* a reward or punishment, or even both.

Previous research has found potential neural correlates/substrates of these systems (Corr, 2008), which we might expect to affect individual differences in behavior. Specifically, individuals with a highly active BAS (i.e. the neural pathways in the areas of the brain related to responding to rewarding stimuli have a low threshold for activation), might exhibit a more positive JB. Conversely, individuals with a highly active FFFS (i.e. highly sensitive to punishment) might exhibit a more negative JB. If both the BAS and the FFFS are highly active, the BIS (the mediator between the two systems) is activated and dogs may display conflict or displacement behavior.

In examining the relationship between Excitability and Fear, it is possible that dogs' responses in the described situations are the result of an increased sensitivity to the distinction between rewarding vs. punishing properties of a stimulus and an exaggerated behavioral response in either case. The RST proposes that in some cases an aversive stimulus should elicit *increased* movement or behavior (e.g. approaching/fighting, or fleeing), while in other cases, it should decrease behavior (e.g. freezing), depending on the proximity and degree of threat that the stimulus presents (McNaughton & Corr, 2008, p. 47). Dogs that are sensitive to external changes may alternate between these responses, and their tendency to become more active vs. less active in any given situation may depend on life history as well as individual differences in temperament. Thus, while they may be correlated in dogs' daily lives, dogs' baseline tendency to freeze versus approach may have had differential impacts in our experimental set up.

Dogs with higher scores on SRB showed a more negative JB than dogs with lower scores in these areas, which is consistent with multiple previous studies examining this relationship

(Barnard et al., 2018; Karagiannis et al., 2015; Mendl et al., 2010b). The reliability of this finding across studies provides strong support for cognitive differences underlying behavior problems in companion dogs. Karagiannis et al.'s (2015) work showed that both JB and the severity of SRB improved with behavior modification and psychopharmacological treatment. These combined results have two major implications: 1) that JB may be an important assessment tool in predicting behavior challenges, and 2) that behavior problems related to a negative JB may require different treatment compared to those related to a more positive JB.

There was an effect of trial order on latency to reach the ambiguous locations, indicating that the dogs learned across the testing trials that the ambiguous locations were not reinforced. This effect was strongest between the first and second presentations of the ambiguous stimuli. Dogs' approach behavior towards the probe stimuli was further affected by whether the trial immediately preceding was a CS+ or a CS-, with a longer latency and decreased probability of approach when the previous trial was a CS+, indicating a more negative JB.

While it may seem counterintuitive that a positive event (receiving a food reward) might induce a more negative JB, it is possible that the recency of a CS+ event improved dogs' discrimination of the probe location, so that their slower rate of approach was based on differentiating the probe location from the previous CS+ trial. Alternatively, Burman et al. (2011) proposed that the anticipation of a rewarding event (procuring food) may induce a more positive JB than is present immediately following a rewarding event. The researchers found that dogs that were given an opportunity to search for food in a maze prior to cognitive bias testing showed a relatively more negative judgement bias than dogs that were taken directly to the testing paradigm, while showing no difference in their behavior towards the reference stimuli. In our paradigm, the recency of a CS+ may have had a similar effect. This effect was mediated by dogs' Trainability score: Dogs whose owners reported that they were more "trainable" showed less of an effect of the preceding trial on their JB towards an ambiguous stimulus.

One notable limitation of our results and interpretation of the relationship between Trainability and dogs' JB in our experiment is that our sample included dogs that were relatively high on Trainability, with a restricted range of scores in that area. In general, dog owners who are willing to volunteer for research may be more interested in dog-centric activities. Our recruitment of owners through local dog training companies may have further exaggerated this tendency, since these owners had already invested in training their dogs. With this in mind, we examined the questions that make up the Trainability scale on the C-BARQ to better understand why it mediated the effect of the previous stimulus on approach behavior in our sample.

The questions used to calculate Trainability scores from the C-BARQ include the frequency of dogs' behavior from "never" to "always" for items such as "When off leash, my dog returns immediately when called," or how often they obey "sit" and "stay" commands immediately. These questions may reflect the amount or success of training owners have already completed, as opposed to how amenable to training the dogs themselves are. Further, the type of training used with dogs (i.e., whether it is primarily reward-based or corrective) may also have an effect on dogs' JB, but because this sample may underrepresent dogs scoring low on Trainability, the overall effect of the Trainability factor is difficult to interpret. Still, given the restriction of range, the diminished effect of the preceding stimulus on dogs' tendency to approach probe stimuli was of particular interest. One possible explanation is that higher Trainability is associated with a decreased behavioral sensitivity to the current emotional state, meaning that dogs' behavior is more consistent across the emotional spectrum.

The location-based discrimination task is not a social interaction, so extrapolating our JB results to dog-dog aggression is the most important limitation of the study. In fact, specific care was taken to minimize handler and owner effects, and testing was conducted with one dog at a time to reduce error and potential confounds. In order to truly understand how judgement bias affects social interactions, and to see if individual differences in judgement bias are consistent across social and non-social contexts, the assessment should use social stimuli. Range, Aust, Steurer, and Huber (2008) demonstrated that dogs can be trained to differentiate pictures of landscapes from images that contain other dogs using an automated touch-screen. Therefore, it may be possible to use a similar paradigm using images of other dogs to see whether dogs classify them as predicting rewards or non-rewards (or punishment). Such a set up would allow assessment of whether the positive judgement bias found in dogs with higher DDA extends to their responses towards conspecific stimuli.

Further, it may be useful to examine what factors are associated with cognitive biases and outcome in social interactions. For example, if body language varies with a dogs' current judgement bias, it may be possible to predict whether a social interaction is likely to go poorly. Likewise, restraint may affect JB in canines, and thereby alter subsequent social interactions. If anticipation of an event (Burman et al., 2011) and release from restraint both induce a positive JB (Doyle et al., 2010), interactions that take place or begin on leash (therefore delaying the actual encounter and extending the period of anticipation) may differ substantially from those in which there is little anticipation and minimal physical restraint.

In summary, our research provides further support that canine behavior, including intraspecific social behavior, is related to individual differences in cognition. Our findings therefore suggest a rich area for future research, as most studies examining factors related to canine aggression are focused on physical differences such as breed variability and neuter status, or on life history such as rearing environment and past training methods. Whereas experimentally modifying these facets may be practically and ethically problematic, past studies have demonstrated that both cognitive ability and cognitive bias can be modified by relatively short-term manipulations (Bateson, Desire, Gartside, & Wright, 2011b; Bethell & Koyama, 2015; Brydges et al., 2011; Douglas et al., 2012; Wichman, Keeling, & Forkman, 2012). Such manipulations offer a promising path forward both for our basic understanding of how cognition impacts social behavior in nonhuman animals, as well as a way of potentially improving the efficacy and efficiency of behavior modification for dogs that show aggression towards other dogs.

4.5 TABLES AND FIGURES

Table 1.
C-BARQ scales and descriptions

Scale	**Description**
Trainability	Responsiveness to owner cues, sensitivity to potential rewards and aversives. Higher score = More “trainable”
Aggression	Severity of aggression displayed in a variety of situations, e.g when verbally or physically corrected, during grooming, etc. Higher score = more severe aggression
Dog-Directed Aggression (DDA)	Severity of aggression displayed during interactions with other dogs. Higher score = more severe aggression
Fear and Anxiety	Severity of fear displayed in a variety of situations, e.g. when hearing a vacuum cleaner or fireworks, or when an unfamiliar human approaches. Higher score = more severe fear
Dog-Directed Fear (DDF)	Severity of fear displayed during interactions with other dogs. Higher score = more severe fear
Separation-related behavior (SRB)	Frequency of particular signs of distress are displayed when the dog is left alone. Higher score = more frequent SRB
Excitability	Degree of reaction to a variety of situations, e.g. when the doorbell rings or before a car ride. Higher score = more excitable
Attachment and Attention-seeking	Frequency with which dog follows or solicits attention from a particular human household member. Higher score = more attachment
Miscellaneous	Frequency of other potential behavior concerns, e.g. eating feces, excessive licking. Higher score = more frequent behaviors.

Figure 1. Average dog age by behavior category. Behavior categories (BehCat) based on owner responses on the C-BARQ. Aggr = Aggressive toward other dogs. Con = generally confident (not fearful). Soc = Sociable with other dogs. Fear = generally fearful. There were no significant differences in average dog age across behavior categories ($F(3,3\ 5) = .0.156$, $p = .925$).

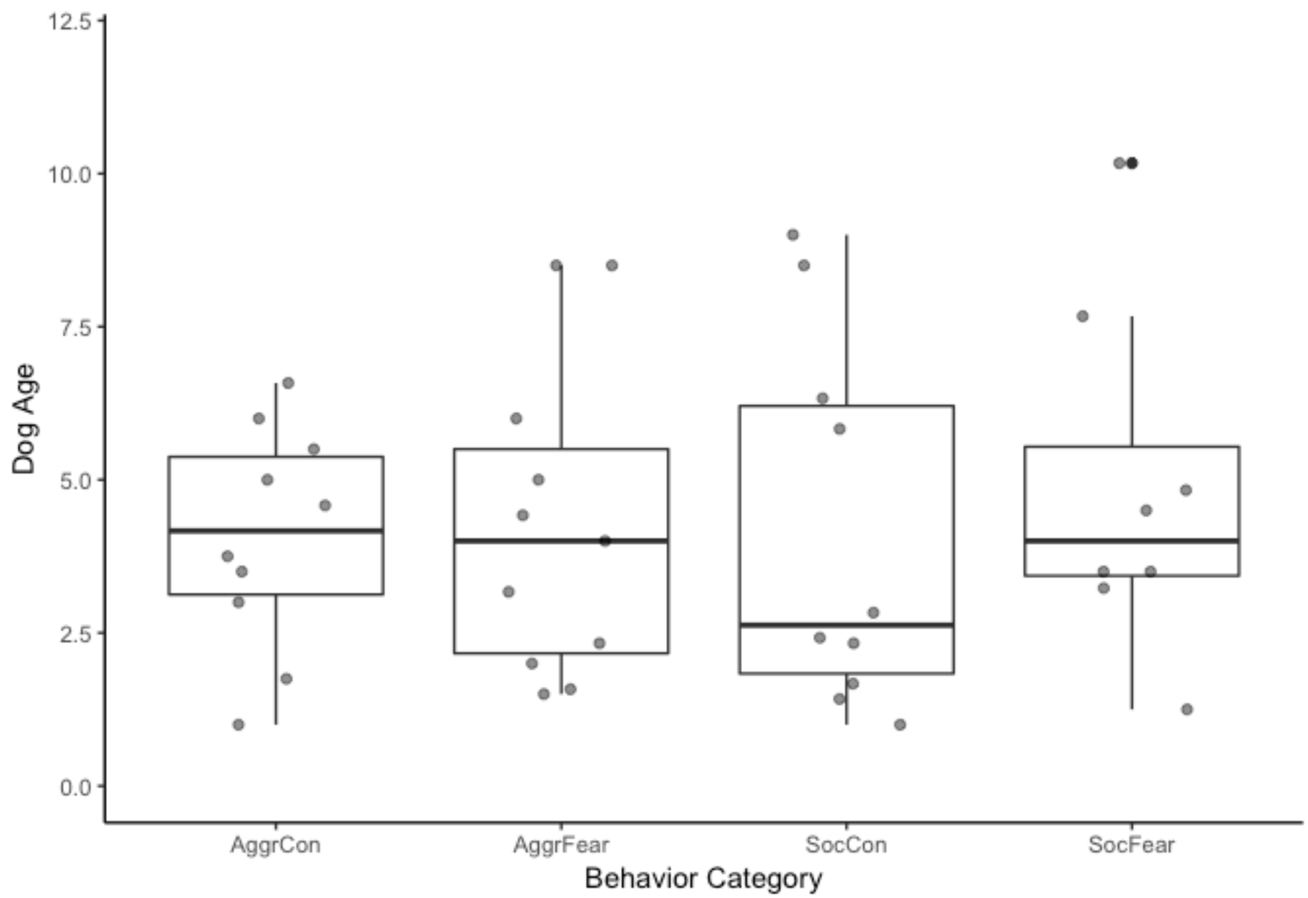


Figure 2. Average duration dogs had been with their owners by behavior category. There were no significant differences in average duration dogs had been owned across behavior categories ($F(3, 35) = 0.992$, $p = .408$). Behavior categories (BehCat) based on owner responses on the C-BARQ. Aggr = Aggressive toward other dogs. Con = generally confident (not fearful). Soc = Sociable with other dogs. Fear = generally fearful.

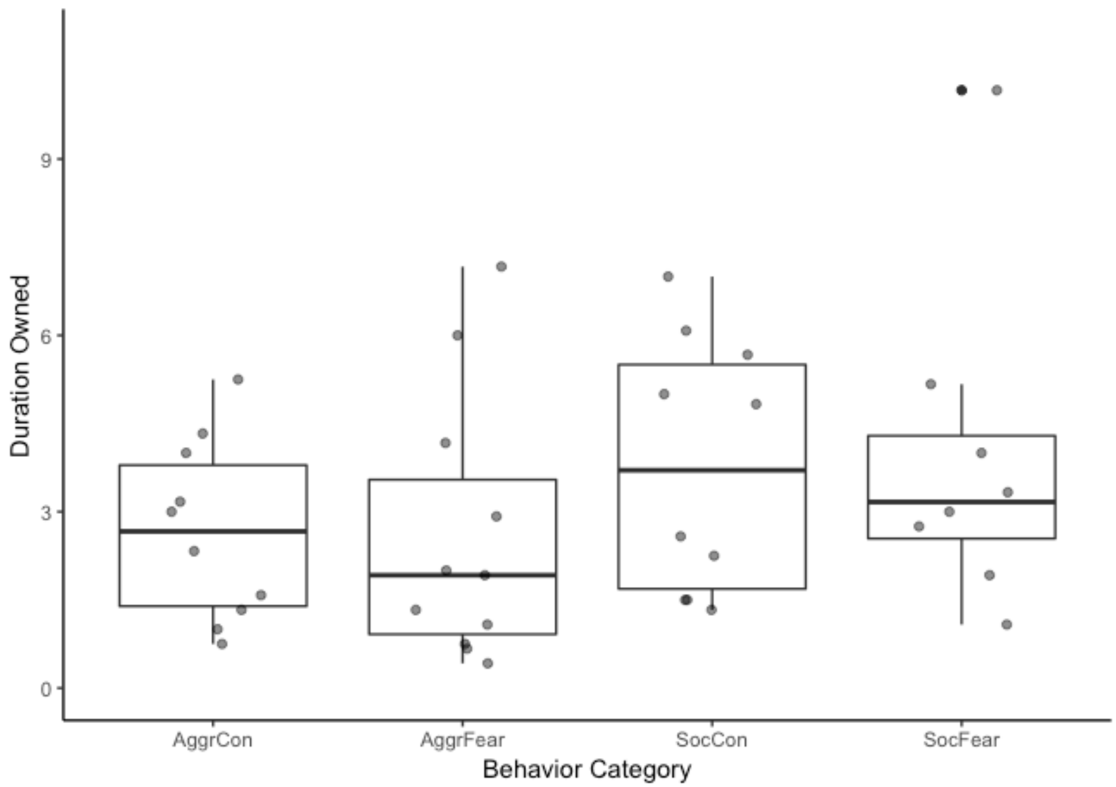


Figure 3. Relationship between number of learning trials to reach criterion and C-BARQ Trainability score. Criterion for learning was that in six consecutive trials, dogs approached the bowl in the CS+ location faster than they approached in any of the CS- trials. Behavior categories (BehCat) based on owner responses on the C-BARQ. Aggr = Aggressive toward other dogs. Con = generally confident (not fearful). Soc = Sociable with other dogs. Fear = generally fearful. Trainability and number of learning trials to reach criterion were not correlated (r = .014, p = .9304).

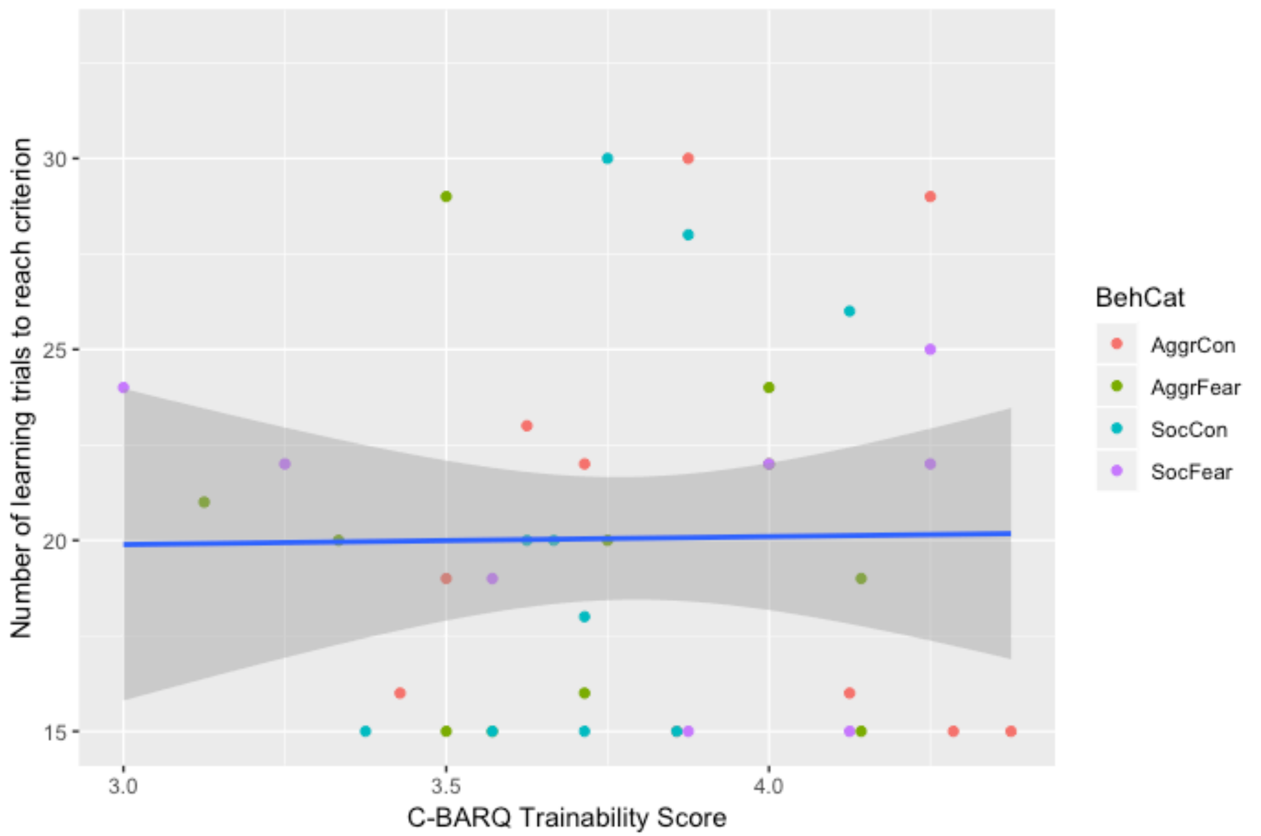


Figure 4. Cox proportional hazards model of dogs' latency and probability of approaching stimuli by location and presentation order. Overall, dogs' approach behavior to the ambiguous stimuli matched their approach to the CS+ more closely than the CS-, demonstrating a positive judgement bias ($z = 7.67$, $p = 1.7 \times 10^{-14}$). They also demonstrated evidence of learning in the probe trials, as they were less likely and slower to approach the bowl in subsequent presentations at the ambiguous locations ($z = 10.43$, $p < .00000001$). This change was significantly greater between the first and second presentations than the second and third presentations ($z = 3.55$, $p = 3.8 \times 10^{-4}$).

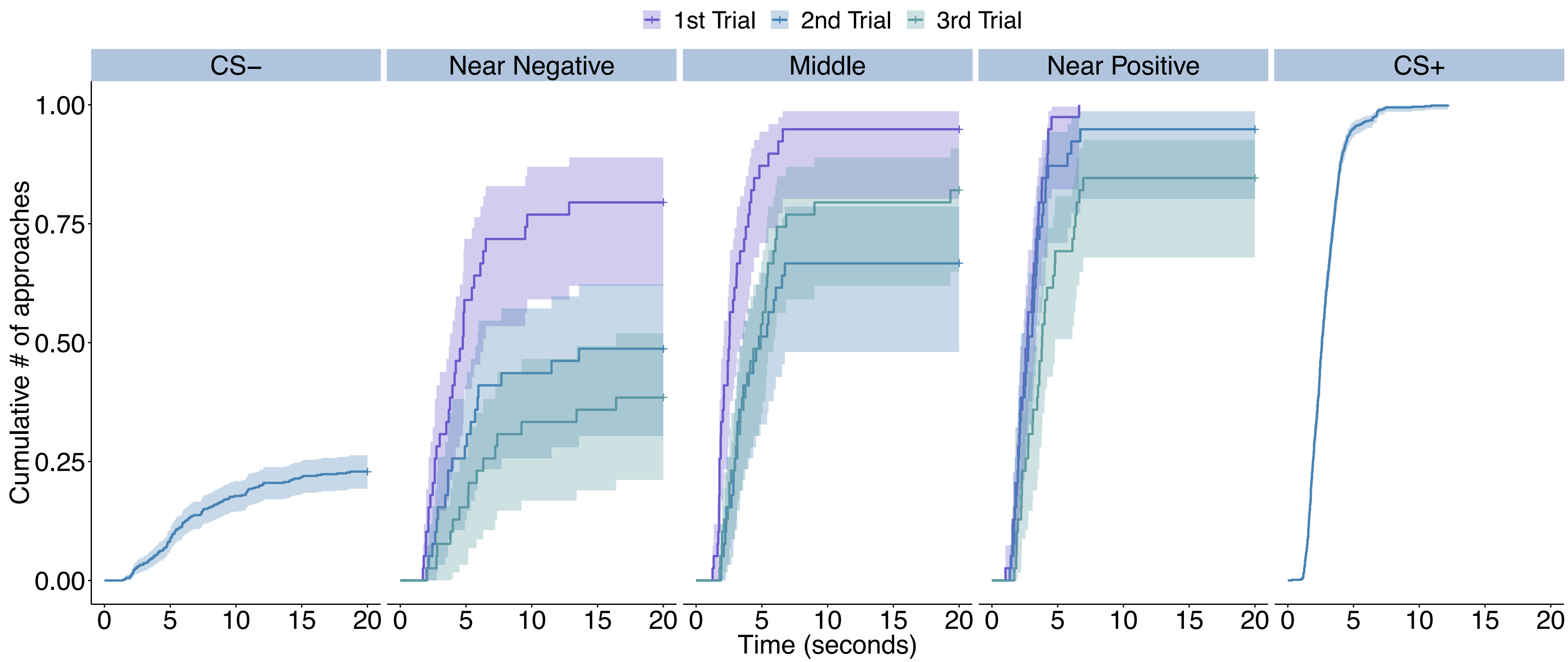


Figure 5. Relationship between C-BARQ Excitability score and latency to approach, by stimulus location. Dogs that were higher in Excitability showed a greater decrease in latency as stimulus locations neared the CS+ ($z = 5.73$, $p = 1.0 \times 10^{-8}$ than dogs that were less excitable.

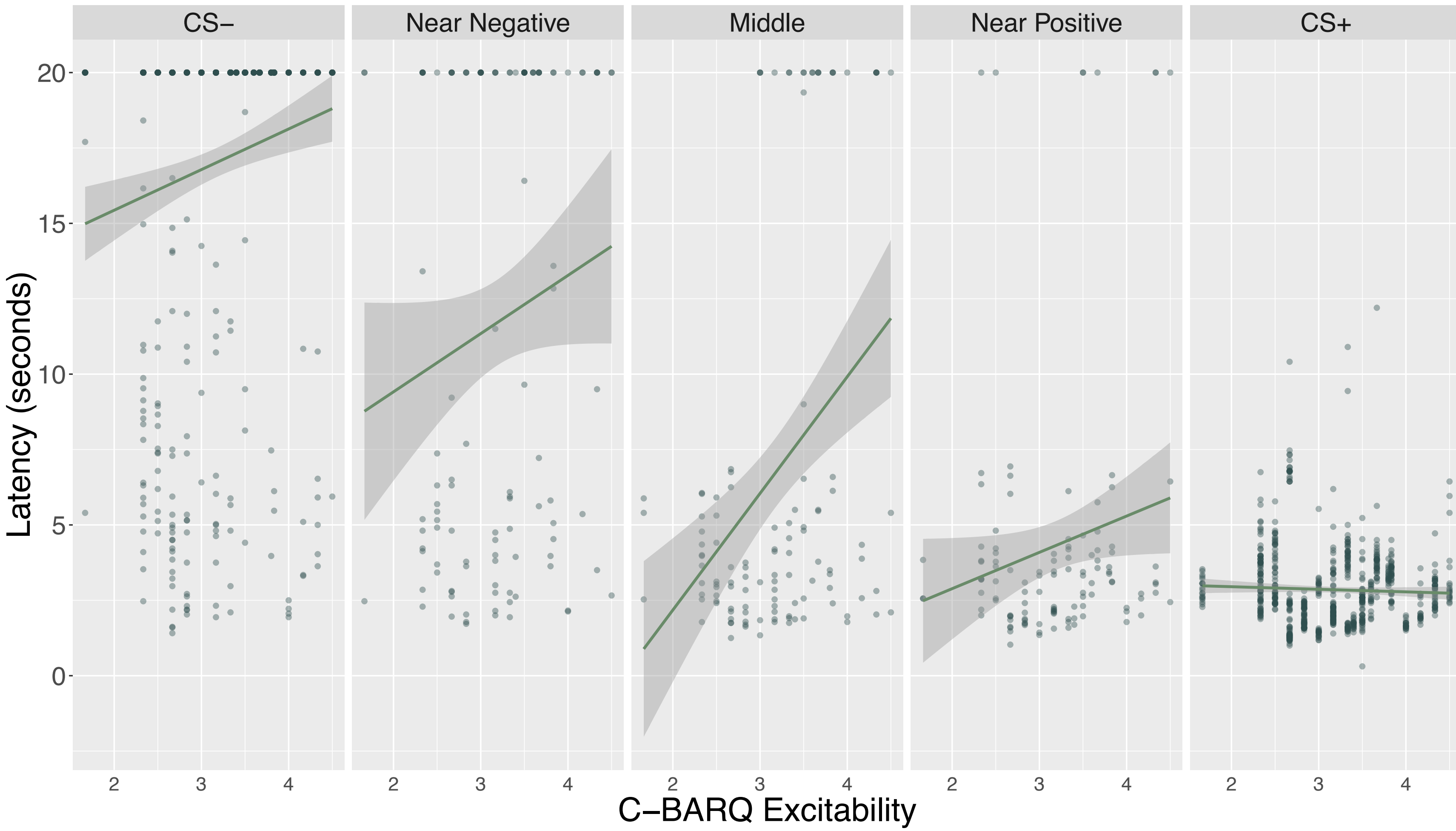


Figure 6. Relationship between C-BARQ SRB score and latency to approach by stimulus location. Dogs with higher scores on SRB showed greater discrimination by stimulus location ($z = 3.36$, $p = 7.9 \times 10^{-4}$), showed greater discrimination between the Near Positive and Near Negative stimuli ($z = 4.20$, $p = 2.7 \times 10^{-5}$) were less likely and slower to approach ambiguous stimuli than dogs with lower SRB scores ($z = 3.25$, $p = 1.2 \times 10^{-3}$), demonstrating a more negative judgement bias as well as increased discrimination of conditioned stimuli.

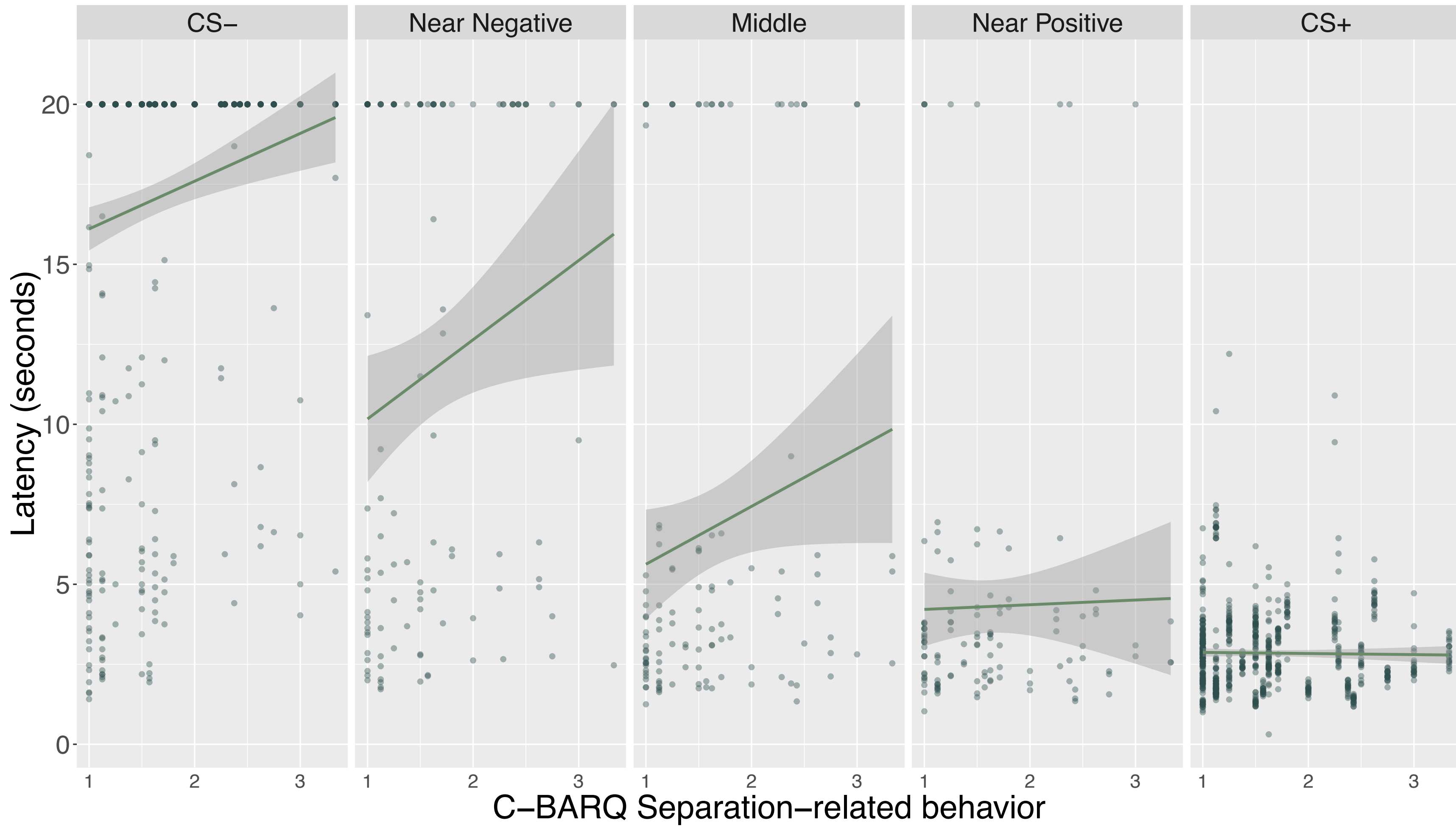


Figure 7. Relationship between C-BARQ Dog-Dog Aggression (DDA) score and latency to approach, by stimulus location. Dogs that were higher in DDA showed a smaller change in speed as the stimulus location moved from CS+ to CS- ($z = 2.29$, $p = 0.02$), and were faster to approach ambiguous stimuli compared to those with lower DDA scores ($z = 3.85$, $p = 1.2 \times 10^{-4}$), demonstrating a more positive judgement bias.

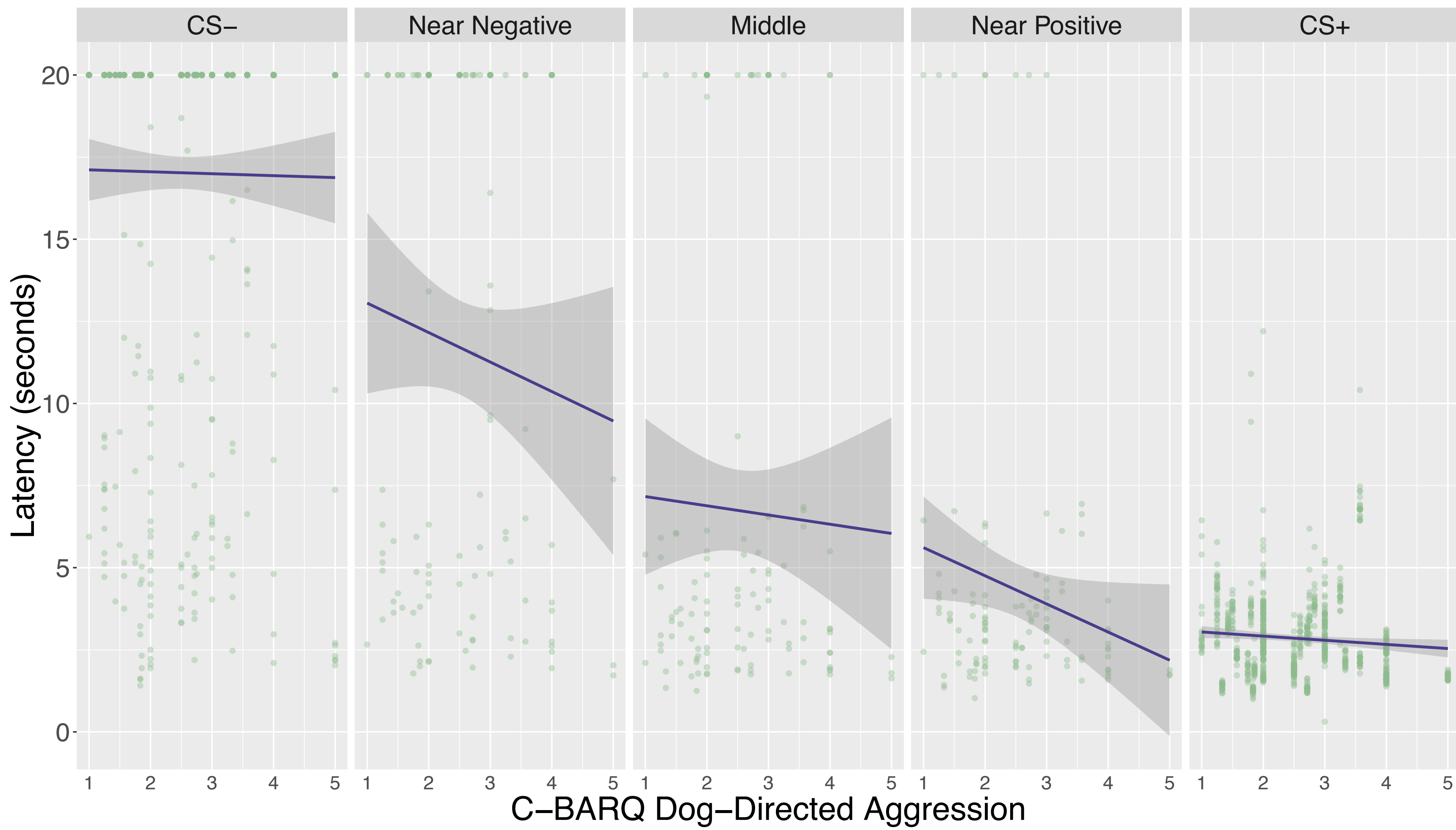


Figure 8. Relationship between C-BARQ Fear score and latency to approach, by stimulus location. Dogs that were higher in Fear showed a more linear change in behavior as probe position moved from CS- to CS+ ($z = 5.71$, $p = 1.1 \times 10^{-8}$), and smaller change in latency between Near Negative and Near Positive probe stimuli ($z = 2.77$, $p = 5.6 \times 10^{-3}$).

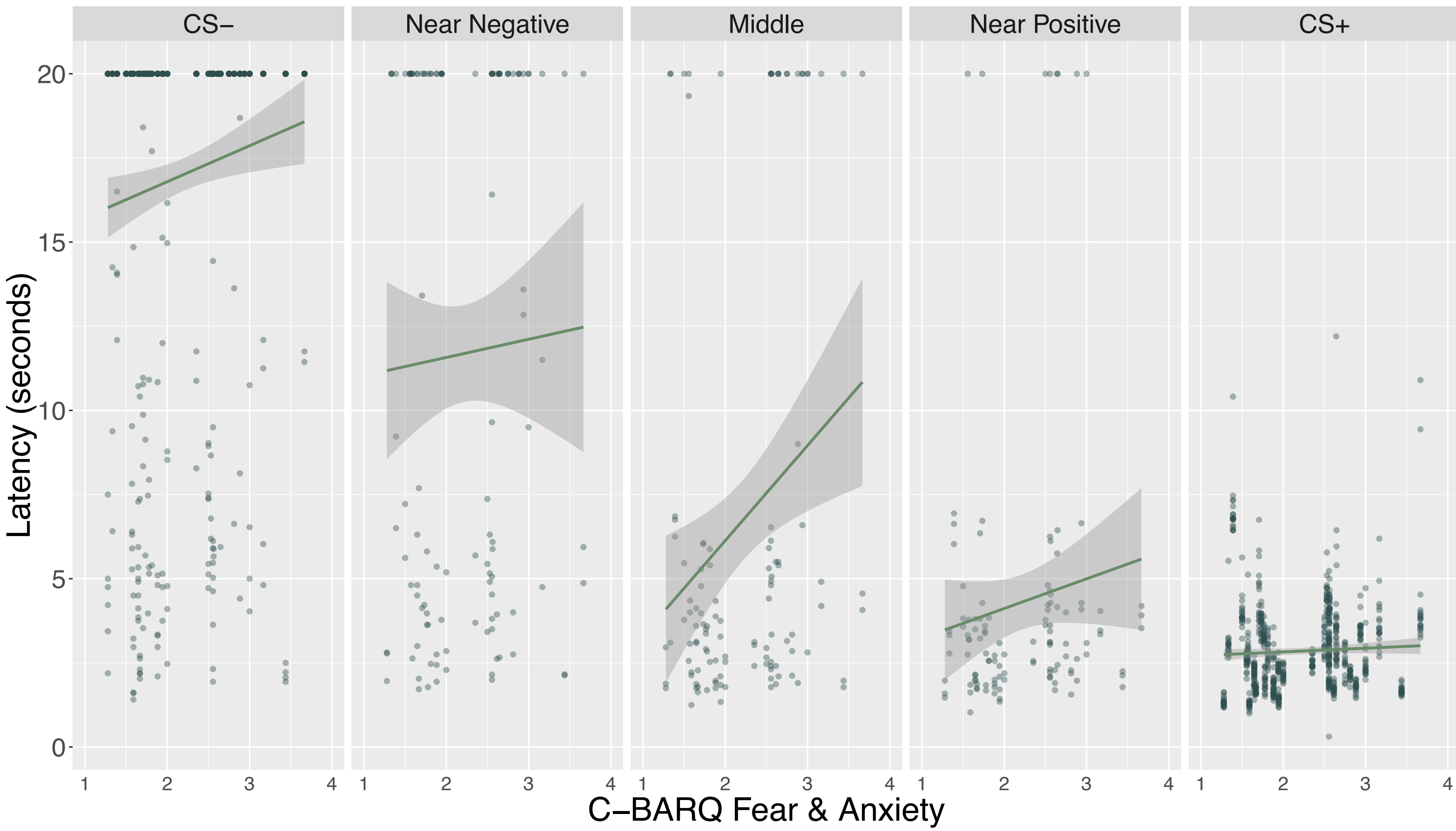


Table 2. Complete Trial order with condition.

Learning Trial #	Condition
1	+
2	+
3	-
4	-
5	+
6	+
7	-
8	+
9	+
10	-
11	-
12	+
13	+
14	-
15	-
16	+
17	+
18	-
19	+
20	+
21	-
22	-
23	+
24	-
25	-
26	+
27	-
28	-
29	+
****30****	+

Refresher trials	Condition
1	+
2	-
3	+
4	-

Testing Trial #	Condition
1	M
2	+
3	-
4	-
5	+
6	NP
7	-
8	+
9	+
10	-
11	NN
12	+
13	+
14	-
15	-
16	NP
17	+
18	-
19	+
20	+
21	NN
22	-
23	+
24	-
25	+
26	M
27	+
28	-
29	+
30	+
31	NN
32	-
33	+
34	+
35	-
36	M
37	-
38	+
39	-
40	+
41	NP
42	+ no treats

4.6 References

Appleby, D. L., Bradshaw, J. W. S., & Casey, R. A. (2002). Relationship between aggressive and avoidance behavior by dogs and their experience in the first six months of life. *The Veterinary Record*, *150*, 434–438.

Barnard, S., Wells, D. L., Hepper, P. G., & Milligan, A. D. S. (2017). Association between lateral bias and personality traits in the domestic dog (Canis familiaris). *Journal of Comparative Psychology*, *131*(3), 246–256. https://doi.org/10.1037/com0000074

Barnard, S., Wells, D. L., Milligan, A. D. S., Arnott, G., & Hepper, P. G. (2018). Personality traits affecting judgement bias task performance in dogs (Canis familiaris). *Scientific Reports*, *8*(1), 1–8. https://doi.org/10.1038/s41598-018-25224-y

Bateson, M., Desire, S., Gartside, S. E., & Wright, G. A. (2011a). Agitated honeybees exhibit pessimistic cognitive biases. *Current Biology*, *21*(12), 1070–1073. https://doi.org/10.1016/j.cub.2011.05.017

Bateson, M., Desire, S., Gartside, S. E., & Wright, G. A. (2011b). Agitated honeybees exhibit pessimistic cognitive biases. *Current Biology*, *21*(12), 1070–1073. https://doi.org/10.1016/j.cub.2011.05.017

Bethell, E. J., & Koyama, N. F. (2015). Happy hamsters? Enrichment induces positive judgement bias for mildly (but not truly) ambiguous cues to reward and punishment in Mesocricetus auratus. *Royal Society Open Science*, *2*(7), 140399. https://doi.org/10.1098/rsos.140399

Borchelt, P. L. (1983). Applied Animal Ethology, 10 (1983) 45-61, *10*, 45–61.

Brydges, N. M., Leach, M., Nicol, K., Wright, R., & Bateson, M. (2011). Environmental enrichment induces optimistic cognitive bias in rats. *Animal Behaviour*, *81*(1), 169–175. https://doi.org/10.1016/j.anbehav.2010.09.030

Burman, O. H. P., Parker, R. M. A., Paul, E. S., & Mendl, M. T. (2009). Anxiety-induced cognitive bias in non-human animals. *Physiology and Behavior*, *98*(3), 345–350. https://doi.org/10.1016/j.physbeh.2009.06.012

Burman, O., McGowan, R., Mendl, M., Norling, Y., Paul, E., Rehn, T., & Keeling, L. (2011). Using judgement bias to measure positive affective state in dogs. *Applied Animal Behaviour Science*, *132*(3–4), 160–168. https://doi.org/10.1016/j.applanim.2011.04.001

Cannas, S., Talamonti, Z., Mazzola, S., Minero, M., Picciolini, A., & Palestrini, C. (2018). Factors associated with dog behavioral problems referred to a behavior clinic. *Journal of Veterinary Behavior*, *24*, 42–47. https://doi.org/10.1016/j.jveb.2017.12.004

Casey, R. A., Loftus, B., Bolster, C., Richards, G. J., & Blackwell, E. J. (2014). Human directed aggression in domestic dogs (Canis familiaris): Occurrence in different contexts and risk factors. *Applied Animal Behaviour Science*, *152*, 52–63. https://doi.org/10.1016/j.applanim.2013.12.003

Corr, P. J. (2008). *The reinforcement sensitivity theory of personality*. (P. J. Corr, Ed.), *The Reinforcement Sensitivity Theory of Personality*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511819384

Douglas, C., Bateson, M., Walsh, C., Bédué, A., & Edwards, S. A. (2012). Environmental enrichment induces optimistic cognitive biases in pigs. *Applied Animal Behaviour Science*, *139*(1–2), 65–73. https://doi.org/10.1016/j.applanim.2012.02.018

Doyle, R. E., Fisher, A. D., Hinch, G. N., Boissy, A., & Lee, C. (2010). Release from restraint generates a positive judgement bias in sheep. *Applied Animal Behaviour Science*, *122*(1), 28–34. https://doi.org/10.1016/j.applanim.2009.11.003

Fatjo, J., Amat, M., Mariotti, V. M., de la Torre, J. L. R., & Manteca, X. (2007). Analysis of 1040 cases of canine aggression in a referral practice in Spain. *Journal of Veterinary Behavior: Clinical Applications and Research*, *2*(5), 158–165. https://doi.org/10.1016/j.jveb.2007.07.008

Goodwin, H., Yiend, J., & Hirsch, C. R. (2017). Generalized Anxiety Disorder, worry and attention to threat: A systematic review. *Clinical Psychology Review*, *54*(May 2016), 107–122. https://doi.org/10.1016/j.cpr.2017.03.006

Hsu, Y., & 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*(9), 1293–1300. https://doi.org/10.2460/javma.2003.223.1293

Karagiannis, C. I., Burman, O. H., & Mills, D. S. (2015). Dogs with separation-related problems show a "less pessimistic" cognitive bias during treatment with fluoxetine (Reconcile™) and a behaviour modification plan. *BMC Veterinary Research*, *11*(1), 80. https://doi.org/10.1186/s12917-015-0373-1

Line, S., & Voith, V. L. (1986). Dominance aggression of dogs towards people: Behavior profile and response to treatment. *Applied Animal Behaviour Science*, *16*(1), 77–83. https://doi.org/10.1016/0168-1591(86)90042-0

McNaughton, N., & Corr, P. J. (2008). The neuropsychology of fear and anxiety: A foundation for Reinforcement Sensitivity Theory. In *The Reinforcement Sensitivity Theory of Personality*. https://doi.org/10.1017/CBO9780511819384.003

Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E., & Casey, R. (2010a). Dogs showing separation-related behaviour exhibit a "pessimistic" cognitive bias. *Current Biology*, *20*(19). https://doi.org/10.1016/j.cub.2010.08.030

Mendl, M., Brooks, J., Basse, C., Burman, O., Paul, E., Blackwell, E., & Casey, R. (2010b). Dogs showing separation-related behaviour exhibit a "pessimistic" cognitive bias. *Current Biology*, *20*(19), R839–R840. https://doi.org/10.1016/j.cub.2010.08.030

Michelangeli, M., Wong, B. B. M., & Chapple, D. G. (2016). It's a trap: Sampling bias due to animal personality is not always inevitable. *Behavioral Ecology*, *27*(1), 62–67. https://doi.org/10.1093/beheco/arv123

Morton, F. B., Lee, P. C., & Buchanan-Smith, H. M. (2013). Taking personality selection bias seriously in animal cognition research: A case study in capuchin monkeys (Sapajus apella). *Animal Cognition*, *16*(4), 677–684. https://doi.org/10.1007/s10071-013-0603-5

Netto, W. J., & Planta, D. J. U. (1997). Behavioural testing for aggression in the domestic dog. *Applied Animal Behaviour Science*, *52*(3–4), 243–263. https://doi.org/10.1016/S0168-1591(96)01126-4

Parker, R. M. A., Paul, E. S., Burman, O. H. P., Browne, W. J., & Mendl, M. (2014). Housing conditions affect rat responses to two types of ambiguity in a reward-reward discrimination cognitive bias task. *Behavioural Brain Research*, *274*, 73–83. https://doi.org/10.1016/j.bbr.2014.07.048

Range, F., Aust, U., Steurer, M., & Huber, L. (2008). Visual categorization of natural stimuli by domestic dogs. *Animal Cognition*, *11*(2), 339–347. https://doi.org/10.1007/s10071-007-0123-2

Sanger, M. E., Doyle, R. E., Hinch, G. N., & Lee, C. (2011). Sheep exhibit a positive judgement bias and stress-induced hyperthermia following shearing. *Applied Animal Behaviour Science*, *131*(3–4), 94–103. https://doi.org/10.1016/j.applanim.2011.02.001

Segrest Purkiss, S. L., Perrewé, P. L., Gillespie, T. L., Mayes, B. T., & Ferris, G. R. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processes*, *101*(2), 152–167. https://doi.org/10.1016/j.obhdp.2006.06.005

Sherman, C. K. C., Reisner, I. R. I. I. R., Taliaferro, L. A., & Houpt, K. A. (1996). Characteristics, treatment, and outcome of 99 cases of aggression between dogs. *Applied Animal Behaviour Science*, *47*(1–2), 91–108. https://doi.org/10.1016/0168-1591(95)01013-0

Tucak, P. (2014). Manual of clinical behavioral medicine for dogs and cats by KL Overall. Elsevier Mosby Publishing, St Louis, 2013. 812 pages. Price: A$79.95. ISBN 9780323008907. *Australian Veterinary Journal*. https://doi.org/10.1111/avj.12192

Verbeek, E., Ferguson, D., Quinquet de Monjour, P., & Lee, C. (2014). Generating positive affective states in sheep: The influence of food rewards and opioid administration. *Applied Animal Behaviour Science*, *154*, 39–47. https://doi.org/10.1016/j.applanim.2014.02.011

Wichman, A., Keeling, L. J., & Forkman, B. (2012). Cognitive bias and anticipatory behaviour of laying hens housed in basic and enriched pens. *Applied Animal Behaviour Science*, *140*(1–2), 62–69. https://doi.org/10.1016/j.applanim.2012.05.006