Increasing arousal enhances inhibitory control in calm but not excitable dogs

Emily E. Bray[a], Evan L. MacLean[b,d], and Brian A. Hare[b,c]

[a]Department of Psychology, University of Pennsylvania, Philadelphia, PA, U.S.A.
[b]Department of Evolutionary Anthropology, Duke University, Durham, NC, U.S.A
[c]Center for Cognitive Neuroscience, Duke University, Durham, NC, U.S.A.
[d]School of Anthropology, University of Arizona, Tucson, AZ, U.S.A.

Corresponding author email: ebray@sas.upenn.edu
Telephone number: 443-257-9072

Abstract

The emotional-reactivity hypothesis proposes that problem-solving abilities can be constrained by temperament, within and across species. One way to test this hypothesis is with the predictions of the Yerkes-Dodson law. The law posits that arousal level, a component of temperament, affects problem solving in an inverted U-shaped relationship: optimal performance is reached at intermediate levels of arousal and impeded by high and low levels. Thus, a powerful test of the emotional-reactivity hypothesis is to compare cognitive performance in dog populations that have been bred and trained based in part on their arousal levels. We therefore compared a group of pet dogs to a group of assistance dogs bred and trained for low arousal (N = 106) on a task of inhibitory control involving a detour response. Consistent with the Yerkes-Dodson law, assistance dogs, which began the test with lower levels of baseline arousal, showed improvements when arousal was artificially increased. In contrast, pet dogs, which began the test with higher levels of baseline arousal, were negatively affected when their arousal was increased. Furthermore, the dogs' baseline levels of arousal, as measured in their rate of tail wagging, differed by population in the expected directions. Low-arousal assistance dogs showed the most inhibition in a detour task when humans eagerly encouraged them while more highly aroused pet dogs performed worst on the same task with strong encouragement. Our findings support the hypothesis that selection on temperament can have important implications for cognitive performance.

Keywords: Inhibitory Control, Arousal, Canine, Cognition, Assistance dogs

Successful problem solving involves well-calibrated emotional and motivational input (Blair and Diamond 2008; Diamond 2010; Hare and Tomasello 2005a; Tooby and Cosmides 2005). This idea is central to the emotional-reactivity hypothesis, which posits that selection on temperament influences problem-solving capabilities in diverse species (Hare and Tomasello 2005a; Hare and Tomasello 2005b). For example, foxes that were selected over generations based on their approach behavior and emotional response to humans are more skilled at using human gestures than a control line bred without regard to their reaction to humans (Hare et al. 2005). The emotional reactivity hypothesis has been proposed to explain shifts in problem solving in a range of taxa including dogs (Hare and Tomasello 2005a), ferrets (Hernádi et al. 2012), bonobos (Hare et al. 2012), and even humans (Cieri et al. 2014).

One largely untested prediction of the emotional reactivity hypothesis is that the effect of temperamental differences on problem solving will be apparent even within species (e.g., Kagan and Snidman 2004; Melis et al. 2006). Dogs provide a particularly powerful test of this prediction given the history of selection that is thought to have focused on temperamental traits such as arousal or excitability (Miklósi 2007). This selection has created a diversity of temperamental profiles that might be explored by comparing subpopulations of dogs on cognitive tasks. Inhibitory control is one problem-solving skill that seems to be affected across taxa by levels of emotional arousal—a component of temperament (Hare et al. 2007; Rosati and Hare 2013; Wright et al. 2011, 2012; Topál et al. 2009)—and is also known to vary widely between individuals and species (MacLean et al. 2014; Moffitt et al. 2011; Bray et al. 2014).

Previous research has shown that the relationship between arousal and problem solving is not always linear. It is theorized that while a higher level of arousal in simple tasks promotes learning, if the task is more cognitively complex increased arousal facilitates performance only to a certain point, beyond which it is detrimental (see Figure 1). Thus, the Yerkes-Dodson (1908) law in its modern interpretation predicts an inverted U-shaped relationship between arousal level and achievement on complex tasks, with performance peaking at moderate arousal levels and suffering at both high and low levels (Duffy 1957; Hebb 1955; Schlosberg 1954). While the predictions of the Yerkes-Dodson law have not always held up (see Watters et al. 1997), there have been a number of instances across species—including humans—in which the U-shaped function between arousal and problem solving has been observed (rats: Broadhurst 1957; chicks: Cole 1911; cats: Dodson 1915; humans: e.g., Anderson 1994; van der Meere et al. 1995). Based on the Yerkes-Dodson law and the complex cognitions involved in exerting inhibitory control, one prediction is that dogs' arousal will affect inhibitory control in an inverted U-shaped curve depending on the temperamental selection that different populations of dogs have undergone.

Pet and assistance dogs, two populations of dogs that vary systematically in level of formal training and artificial selection, provide one way to test this prediction. Pet dogs are attuned to human gestures, but generally receive no professional training (beyond basic obedience) or systematic genetic selection (other than that which occurs spontaneously in pet dog populations). Assistance dogs, on the other hand, may be hyper-attuned to human gestures as a result of both intensive training and intentional, highly-controlled selection (e.g., Topál et al. 2006). Furthermore, as part of their training, many working dogs are required to perform acts that require inhibitory control, such as following commands while a cat walks around the training area or while ignoring scattered dog food. Failure on tests of inhibitory control have been linked to high aggression, decreased tolerance of close contact, and negative responses to novelty in dogs (Wright et al. 2011). All of these traits are highly discouraged in assistance dogs and could lead to release from training and/or breeding programs.

We predicted that there would be measurable differences between pet and assistance dog populations in how arousal affected their ability to exercise inhibitory control and that each dog type would be impaired by either under- or over-arousal. Specifically, assistance dogs tend to have placid temperaments as a result of selective breeding for these characteristics and extensive training related to self-regulation in the face of arousal or distraction. In the absence of such purposeful selection and training, pet dogs as a whole tend to be more temperamentally reactive than assistance dogs. Thus, we expected pet dogs to be more prone to errors due to over-arousal, whereas assistance dogs might be more prone to errors due to under-arousal.

To explore variation in inhibitory control among dogs in different arousal contexts, we tested 30 pet dogs and 76 candidate assistance dogs. Dogs were required to detour a fence to retrieve a reward, temporarily creating distance between themselves and the food. Each dog participated in both high and low arousal trials. In the high arousal trials the experimenter called the dog in an urgent, high-pitched tone of voice whereas in the low arousal trials she used a low, monotone voice.

Recruitment & Owner Consent

Pet dogs were recruited through and tested at the Duke Canine Cognition Center (DCCC). Owners from the Raleigh-Durham, North Carolina area completed a Dog Registration Questionnaire (http://bit.ly/AmWURq) on the DCCC website in order to be added to a database, which was then screened to remove dogs with histories of aggression and/or restrictive health issues. While some dogs had visited the DCCC up to three previous times, they were all naïve to the testing apparatus and procedures. Owner and dog participation was voluntary, and all owners signed informed-consent forms prior to beginning the experiment.

Assistance dogs were in training at Canine Companions for Independence (Santa Rosa, CA), a national non-profit organization that provides assistance dogs to people with disabilities. Forty-six dogs participated in a four-day cognitive test battery which included this experiment. Thirty additional dogs were recruited solely for this experiment. All dogs were either in their 1^{st}, 2^{nd}, or 3^{rd} (four month-long) semester of training, and naïve to the testing apparatus and procedures.

All testing procedures adhered to regulations set forth by the Duke Institutional Animal Care and Use Committee (IACUC # 303-11-12).

Subjects

Forty pet dogs came to the Duke Canine Cognition Center to be tested, but 10 of these dogs were unable to complete testing. Based on preset abort criteria, a dog was excluded if she had to repeat any one trial more than four times, had to repeat a total of eight trials over the entire session, or did not eat food within 30 s when the food was placed directly in front of her. If any of these *a priori* conditions were met, the session was aborted, and partial data from these sessions were excluded from analysis. The 10 dogs that did not finish the testing session were unable to do so for a variety of reasons (see Online Resource 1). Thus, 30 pet dogs, 16 male and 14 female (mean age = 62.15 months; range 7.8 - 137.1 months) were included in this study. In addition, 77 assistance dogs were tested at their training center, but one of these dogs was unable to complete testing (see Online Resource 1). In total, 76 assistance dogs, 29 male and 47 female (mean age = 25.35 months; range 19.6 – 31.3 months), participated in this study. See Table 1 and 2 for a list of subjects' breeds, sexes, and ages.

Methods

We tested dogs with different training backgrounds to see how they differed on a task that varied arousal level. In this experiment, dogs were presented with a detour task which required inhibitory control because, while subjects could see the food close by, to gain the reward they first had to walk around the transparent barrier, temporarily increasing the distance between themselves and the reward. This type of task has been shown to present an inhibitory challenge for many dogs (e.g., Frank and Frank 1982; Pongrácz et al. 2001; Marshall-Pescini et al. 2015; Osthaus et al. 2010). Furthermore, it is very similar in design and demands to a detour reaching task, shown to be indicative of prefrontal-dependent response inhibition by a rich cross-species body of research (e.g. Humans: Diamond 1990; Macaques: Diamond et al. 1989; Squirrel monkeys: Parker et al. 2005; Apes: Amici et al. 2008; Vlamings et al. 2010; Song sparrows: Boogert et al. 2011). In fact, consistent individual differences in performance on this detour reaching task are observed in squirrel monkeys based on stress-inoculation (exposure to mild stress) early in life, and remain stable up to three and a half years later (Parker et al. 2012).

To assess the role of arousal on the problem-solving skills of assistance and pet dogs, we used a within-subject design in which each dog experienced a series of both low and high arousal trials. In the low arousal trials, the experimenter called the dog in a calm, monotone voice, while in the high arousal trials, the experimenter called the dog in a high-pitched, excited voice.

Apparatus

Two garment racks with transparent shower curtains were placed in a v-shaped fence formation opening away from the dog. Each of the two side panels was approximately 80 cm wide and 1 m tall. The experimenter stood on the opposite side of the curtain, visible behind a sheet of transparent shower curtain that was approximately 40 cm wide and 1 m tall (Figure 2a, b). The dog-handler centered the dog approximately 1.5 m from the front of the apparatus at the start of each trial. Treats were Real Meat® Jerky in beef, chicken, venison, or fish & venison flavors. For the assistance dogs and some pet dogs, the food treats were also paired with a Kong. All sessions were video-recorded.

Procedure and Design

Methods were adapted from a prior study on spatial navigation (Pongrácz et al. 2001). All dogs completed a familiarization trial followed by a block of five “low arousal” trials and a block of five “high arousal” trials. The ordering of the blocks was counterbalanced across dogs: 15 pet dogs and 46 assistance dogs received high arousal trials followed by low arousal trials (order A), while the other 15 pet dogs and 30 assistance dogs received low arousal trials followed by high arousal trials (order B). The assistance dogs could not be completely counterbalanced because 46 of the dogs participated in this test as part of another long-term study which required them all to complete the task in the same order. In both orders, dogs received a two-minute break between these trial blocks, during which time the dog was petted and calmed. The dogs were not given treats during this interval.

Familiarization trial. The handler walked the dog, on lead, completely around the entire perimeter of the apparatus. This trial ensured that the dog had experience maneuvering around the apparatus and acquired knowledge of the motor response required during the test trials.

Low Arousal trials. At the start of each trial, the handler centered the dog at the start line and the experimenter showed the dog the treat that she was holding. The experimenter then crouched behind the fence and vocalized toward the dog in a low, monotone voice. She said “[Dog’s name], look, [Dog’s name], look” during this time. After three seconds elapsed, the handler dropped the leash and the dog was allowed to move toward the experimenter, who continued to vocalize, now saying “[Dog’s name], come” (see Online Resource 2). At no point during the trial did the handler ever prompt or vocalize toward the dog.

A trial was repeated if the dog did not make any responses (defined as either an ‘around’ or a ‘front’ response—see scoring and analysis section) within 20 s. If a dog repeated a single trial four times or had to repeat eight trials over the course of the session, she was excluded from the study (see Online Resource 1). On every trial, the handler started a stopwatch when the experimenter began vocalizing and stopped it when the dog retrieved the food. All trials had a maximum duration of two minutes—thus, if the dog made a response (as defined below) within 20 s but was unable to solve the problem within two minutes, the dog received the maximum latency of two minutes and the handler then walked the dog around the apparatus to receive the treat from the experimenter.

High Arousal trials. High arousal trials were identical to low arousal trials, except that rather than speaking in a monotone voice, the experimenter addressed the dog in a high-pitched, excited voice (See Online Resource 2). The experimenter also enthusiastically waved the treat back and forth and made large arm movements. Because the reward was never hidden from the dog, no attempts were made to control for odor cues throughout the task.

Scoring and Analysis

First, as a measure of each dog’s arousal level before and during the task, we coded tail-wagging rates (in wags per minute) from video. Past studies have used tail-wagging levels as one measure of both positive and negative arousal level (e.g., Freedman et al. 1961; Rehn and Keeling 2011; Pluijmakers et al. 2010; McGowan et al. 2014; Prescott et al. 2004).

Familiarization tail-wagging rates. Tail-wagging rate was coded for each dog during the familiarization trial, which began with the walk around the apparatus and ended at the experimenter's first command at the start of the first test trial. If the dog disappeared from view at any point, that amount of time was excluded. A tail wag was operationalized as the tail moving back and forth (e.g., left to right) horizontally once.

Test tail-wagging rates. Tail-wagging rate was coded throughout each high arousal trial and each low arousal trial, from the moment the handler dropped the leash, until the dog successfully retrieved the treat. The same criteria as above were applied.

Next, as an indicator of performance, two measures of accuracy ("touch" and "pathway") and one measure of latency ("time to success") were coded from video by the primary experimenter (EB).

Touch. The experimenter recorded whether or not the dog touched the barrier (1/0, a binary measure). A touch was coded if the dog's muzzle, nose, or forepaw made physical contact with the outside of the shower curtain. The touch could be directed to either the front panel of the shower curtain or either of the side panels of the shower curtain. If the dogs were tempted to approach the reward directly, making contact with the barrier was seen as representing an inhibitory failure.

Pathway. The experimenter also recorded whether the dogs' initial approach was toward the front of the apparatus (coded as a "front" response) or around the side (coded as an "around" response). A "front" response occurred when the dog came to within 18 in of the front of the apparatus, an area that was marked on the floor. The front response was assumed to represent a lack of inhibitory control, while the around response was assumed to indicate that the dog inhibited its tendency to approach the food directly, choosing instead to take the more circuitous, but effective, route around the barrier.

Time to success. The handler recorded the time, to the nearest tenth of a second, that it took for the dog to complete the detour and retrieve the reward on each trial. The handler timed each trial with a stopwatch. The timing began with the dog's first step forward after the handler released the leash, and ended when the dog retrieved the reward from the experimenter. Longer latencies to complete the task were assumed to designate worse inhibitory control, as dogs that were distracted by the treat would make time-consuming perseverative errors (See Online Resource 2).

All measures were coded from video by the primary experimenter (EB), using a stopwatch for the time measures. Two camera angles were used for coding: one camera with a wide-angle lens was positioned in the back corner of the room behind the start line, so that the dog, handler, apparatus, and experimenter were in view, allowing time to success and tail wagging during the entire trial to be coded. The second camera was positioned on the side of the apparatus and zoomed in, so that each dog's choices and the experimenter were visible, allowing for up-close views of each dog's pathway and touch measures in particular. Twenty percent of trials were randomly selected and coded from video by a second individual who did not participate in the experiment and was naïve to the hypotheses. In terms of arousal measures, the inter observer reliability for pet dogs was excellent for familiarization tail-wagging rates ($r_s(4) = 0.93$, $p < 0.001$) and very good for test tail-wagging rates ($r_s(58)= 0.85$, $p < 0.001$). The inter observer reliability for assistance dogs was very good for familiarization tail-wagging rates ($r_s(13) = 0.86$, $p < 0.001$) and good for test tail-wagging rates ($r_s(148) = 0.76$, $p < 0.001$). In terms of performance measures, the inter observer reliability for pet dogs was very good for pathway (kappa = 0.86) and touch (kappa = 0.88) and excellent for time to success ($r_s(58) = 0.95$, $p < 0.001$). The inter observer reliability for assistance dogs was good for pathway (kappa =

0.74) and excellent for touch (kappa = 0.97) and time to success ($r_s(148) = 0.97$, $p < 0.001$). In cases of disagreement, the original coder's measures were used.

All data were analyzed using R statistical software (version 3.1.1, R Foundation for Statistical Computing, R Development Core Team, 2009). All tests were two-tailed.

Results

A two-way repeated-measures analysis of variance on test tail-wagging rates showed a significant effect of trial type ($F_{1,104} = 195.76$, $p < 0.001$). Dogs wagged their tails more frequently during high arousal than low arousal trials indicating that the experimental manipulation did indeed affect subjects' arousal levels (High arousal: assistance M = 124.76 wags/min, SD = 28.09 wags/min, pet M = 122.06 wags/min, SD = 48.50 wags/min; Low arousal: assistance M = 92.39 wags/min, SD = 29.75 wags/min, pet M = 96.60 wags/min, SD = 47.21wags/min). There was no significant main effect of population (pet vs. assistance; $F_{1,104} = 0.01$, $p = 0.92$) and no significant interaction between trial type and population ($F_{1,104} = 2.05$, $p = 0.16$). As a further test that experimenter arousal affected dog arousal, one-tailed binomial tests indicated that 95% of assistance dogs and 90% of pet dogs showed higher average tail-wagging rates during high arousal trials than low, which is significantly greater than the amount that would be expected by chance, $p < 0.001$.

There was a significant difference between populations, $t_{35.22} = -3.26$, $p = 0.002$, with assistance dogs wagging their tails less rapidly (mean rate = 36.37 ± 3.13 wags/min) than pet dogs (mean rate = 69.54 ± 9.68 wags/min) prior to the test. Furthermore, there were no significant differences within the two orders of assistance dogs, $t_{57.03} = 1.23$, $p = 0.22$, or the two orders of pet dogs, $t_{26.23} = 0.43$, $p = 0.67$; assistance dogs that experienced the high arousal first order were not significantly different in their familiarization tail-wagging rates (mean rate =

33.19 ± 3.82 wags/min) than assistance dogs that experienced the low arousal first order (mean rate = 41.23 ± 5.29 wags/min), and the same was true of pet dogs in the low arousal first (mean rate = 65.32 ± 11.95 wags/min) and high arousal first (mean rate = 73.76 ± 15.58 wags/min) orders. Therefore, these data support the hypothesis that the two populations began the test at differing levels of arousal, with assistance dogs beginning the test with lower baseline arousal levels than pet dogs.

The three performance measures that reflected inhibitory control—touch, pathway, and time to success)—were strongly positively associated with one another. A chi-square test of independence between touch and pathway was significant [$X^2_{(1,\ N = 1060)} = 356.90$, $p < 0.001$], revealing that dogs that followed an “around” pathway were significantly less likely to touch the apparatus. A linear mixed-effects model with dog ID as a random effect, time to success as the dependent variable, and pathway and touch as the predictor variables showed that pathway ($F = 93.84$, $p < 0.001$) and touch ($F = 72.40$, $p < 0.001$) were both significant predictors of a dog’s time to success. We therefore combined these measures into a single composite measure of performance, giving equal weight to each measure. Each dog’s composite score on each trial was defined as the sum of her score on: touch (0 = no touch or 1 = touch), pathway (0 = around pathway or 1 = front pathway), and time to success (0 through 120 seconds). Since trials were capped at 120 seconds, we took each “time to success” score, which was originally recorded in seconds, and divided it by 120, meaning the scores would now fall between 0 and 1 (where the fastest time = 0.0 and the slowest time = 1.0). Across all three individual measures lower scores indicated more successful behavior, and so lower composite scores corresponded with better performance. In pet dogs, the composite response scores ranged from 0.015 to 3 with a mean and

SEM of 0.70 ± 0.047 and in assistance dogs, the composite response scores ranged from 0.013 to 3 with a mean and SEM of 0.41 ± 0.026.

With the composite response score as the dependent variable, we used a linear mixed model with trial type (low arousal vs. high arousal), order (low arousal first vs. high arousal first), trial number (1-10), and population (pet vs. assistance) as fixed effects, and dog ID as a random effect. We also included two interactions, population by trial type and population by order, to investigate the possibility that the problem solving of assistance and pet dogs is affected differently by arousal level.

We first performed a likelihood ratio test to compare the linear mixed model with all predictor variables and the two interactions as predictors (the full model) against a null model (Crawley 2005). The full model fit the data significantly better than the null model (X^2= 252.9, df = 6, p < 0.001).

The main findings are summarized in Table 3 and Figure 3, which shows results separately for assistance and pet dogs. The full model revealed a significant main effect of trial number; almost all dogs improved (that is, achieved lower composite response scores) over time. There was also a significant interaction between population (pet, assistance) and trial type (Figure 4a). (Separately designating each of the three outcome variables that made up the composite as the sole outcome measure in a linear mixed model returned the same results as the model reported here, with the trial type by population interaction and pattern holding in all models [Touch model: t = -3.15, p = 0.002; Pathway model: t = -5.34, p < 0.001; Time to success model: t = 1.56, p = 0.007]). Therefore we used contrasts to investigate the subgroup-specific effects of trial type—i.e., the effects within assistance and pet dogs. These analyses revealed that assistance dogs performed significantly better in high arousal than low arousal trials (b = -0.28, z

= -6.35, $p < 0.001$). In contrast, pet dogs achieved significantly better composite scores during low arousal than high arousal trials ($b = 0.42$, z = 6.18, $p < 0.001$). Thus, while the trial type influenced performance in both populations, it had opposite effects between pet and assistance dogs.

Additionally, there was a significant interaction between population and the order in which high and low arousal trials were administered (Figure 4b). Contrasts revealed that assistance dogs achieved significantly better composite scores when facing the block of high arousal trials first ($b = -0.25$, $z = -2.91$, $p < 0.01$). In contrast, pet dogs achieved better composite scores when facing the block of low arousal trials first, although the effect of order was not significant for pet dogs ($b = 0.22$, $z = 1.65$, $p = 0.099$).

Finally, our sample of pet dogs included some dogs that were smaller and some that were older than in our sample of assistance dogs. To rule out the possibility that these size or age differences were driving the effects we observed, we removed the smallest third (n=9 excluded, all under 35 pounds) of pet dogs from the model and found that the results did not change (see Online Resource 3). We then removed the oldest third (n=10 excluded, all over 74 months) of pet dogs from the model and again found similar results (see Online Resource 3). Furthermore, pet baseline arousal levels as measured by tail-wagging rates were not significantly correlated with size (r = -0.34, $p = 0.07$, n = 29) or age (r = 0.07, $p = 0.73$, n = 30).

Our results indicate that pet dogs benefit significantly from low-arousal scenarios, presumably due to their naturally higher levels of baseline arousal as a group, while assistance dogs benefit from high-arousal scenarios, presumably due to their naturally mild levels of baseline arousal as a group. While these results are consistent with the Yerkes-Dodson hypothesis, it would be ideal to test a third population along the continuum—i.e., a group of dogs that has a medium level of arousal. The prediction in this case would be that this group would be least affected by the manipulation from low to high arousal. However, since it is not immediately intuitive which group of dogs would fall between pet and assistance dogs in terms of arousal, we instead approached the problem by momentarily disregarding dog population membership (i.e., pet versus assistance) and instead grouped dogs by their baseline arousal at the start of the task, as measured by tail-wagging rate.

In order to investigate different points along the continuum of arousal in our data, we used baseline tail-wagging rate to split all of the dogs into percentiles and then looked at the lowest, middle, and highest groups. The lowest arousal group consisted of dogs in the first quintile (n=23), the middle arousal group consisted of those in the third quintile (n=21), and the highest arousal group consisted of those in the fifth quintile (n=22). We then assigned each dog a difference score, comprised of their composite score on the high arousal trials minus their composite score on the low arousal trials. Here, negative scores indicated better performance under high arousal conditions, scores close to zero indicated no strong difference between conditions, and scores above zero indicated better performance in low arousal conditions versus high. By plotting the average difference scores of each group, we observed a pattern that although not significant, is consistent with the U-shaped function predicted by the Yerkes-Dodson law (Figure 5). In other words, dogs in the lowest arousal group benefited most from increased arousal, dogs in the middle did not differ much between trial types, and dogs in the highest baseline arousal group suffered most from increased arousal. Thus, with respect to the curve, dogs starting with low arousal move toward optimal while dogs starting with high arousal are pushed further away from optimal, with dogs in the middle being the least affected.

Consistent with our hypothesis that pet dogs generally have the highest arousal and assistance dogs have the lowest, the skew of the groups aligned as we would expect: the low and middle arousal groups consisted of predominantly assistance dogs (low: 18 assistance, 3 pet; middle: 18 assistance, 5 pet), while the high arousal group was composed of mostly pet dogs (high: 12 pet, 10 assistance).

Discussion

The results of the current study provide further support for a link between emotional reactivity and cognitive performance. Temperament not only plays a role in cognitive performance across species, but within a species as well. Applying the Yerkes-Dodson law to the current experiment, we predicted that 1) pet dogs would have higher baseline levels of arousal than assistance dogs and 2) inducing arousal would negatively affect inhibitory control in pet dogs while enhancing it in assistance dogs. We found that assistance and pet dogs differed in their baseline arousal levels when assessing their relative tail-wagging rates. An experimenter was also able to manipulate the dogs' arousal using excited vocal prompts since in both populations tail wagging increased as a result. Finally, assistance dogs with low baseline arousal showed an improvement in performance on the detour task with increased arousal while pet dogs that had relatively higher basal arousal levels showed the opposite pattern. These results suggest that high arousal trials hindered the performance of pet dogs while bolstering the performance of assistance dogs. One explanation for these findings derives from the Yerkes-Dodson (1908) law, which in its modern from posits that arousal and performance on a cognitively complex task follow an inverted U-shaped function, in which optimal performance is reached at an intermediate level of arousal with under- and over-arousal harming performance (also see Dodson 1917; Hebb 1955).

Overall, both populations benefited through increased experience with the task, as evidenced by their improving composite scores over the course of the session. Additionally, arousal state is a powerful predictor of how well a dog will solve this detour problem, but the two groups noticeably differed in the way in which arousal affected their problem-solving success. We attribute the differential performance to dissimilarities in temperament arising from differences in the training and rearing history of pet and assistance dogs.

Alternatively, these temperamental differences could be due to a dog's size or age, with small and/or young dogs being more excitable. Indeed, studies have shown an inverse correlation between hyperactivity/excitability and body size in dogs (McGreevy et al. 2013; Serpell and Duffy 2014). In our study, assistance dogs were relatively homogenous with respect to both factors, while pet dogs were more variable: there were greater numbers of old and small pet dogs as compared to assistance dogs. Thus, one possibility is that the smaller pet dogs drove this effect, and the key difference between the two groups was not pet versus assistance dogs per se, but rather small versus large dogs. However, our data do not support this hypothesis: when we removed the smallest or oldest third of pet dogs from the model, our results did not change. Thus, while age and size probably do play a role in temperament, neither factor is sufficient to explain our results.

Previous studies investigating the links between temperament and cognition in nonhuman animals have found that emotional reactivity is linked to outcomes in social problem-solving tasks (e.g., Hare et al. 2005; Hare et al. 2007). For example, bonobos are more behaviorally tolerant of one another, and can thus solve some cooperative problems with less constraints than chimpanzees (Hare et al. 2007; MacLean and Hare 2013). Another study of ape and monkey species found that the best predictor of inhibitory control was whether or not the animal belonged to a species characterized by high fission-fusion dynamics, suggesting that evolving in a social environment that promotes behavioral flexibility can positively impact such cognitive skills (Amici et al. 2008).

Here we have used two populations of dogs to demonstrate a nuanced, within-species effect, wherein each population's baseline arousal state interacted with experimentally induced changes in arousal, in a manner consistent with the Yerkes-Dodson law. Because the two populations began the experiment with different baseline states of arousal, these conditions allowed pet dogs to perform "better" in one context and assistance dogs to perform "better" in the other. These results have important implications for how we understand cognitive evolution. Namely, selection for specific temperamental profiles may lead to species level differences in problem solving that are moderated by the conditions under which a species is tested. Specifically, populations or species with low baseline states of arousal may perform optimally under states of heightened arousal whereas the opposite would be predicted for species with higher states of baseline arousal.

Our results can be compared to human studies that administer caffeine to manipulate physiological arousal and report an inverted U-shaped function between performance and arousal, consistent with the Yerkes-Dodson hypothesis (e.g. Anderson 1994; Revelle and Loftus 1992; Anderson 1990). Researchers hypothesized that the observed parabolic relationship between performance and arousal crucially hinges on a third factor: personality of the individual subjects (e.g., Revelle and Loftus 1992; Broadhurst 1959), and specifically "arousability" (Eysenck 2002). Those who are chronically at higher levels of arousal become over-aroused and perform poorly in high arousal scenarios, whereas those who are chronically at lower levels of arousal perform best in the same situations, and vice versa.

Future work should address the extent to which these differences can be attributed to training and rearing factors vs. innate genetic differences between populations. For example, it would be informative to compare this population of assistance dogs to assistance dogs that were not specifically bred for working roles, but which have undergone a similarly rigorous training program. Furthermore, future research will benefit by including additional measures of individual differences in temperament as predictors of problem-solving abilities. While we used prior training histories and tail-wagging rates as proxies for temperament, temperamental traits could also be measured using physiological parameters (e.g. heart rate variability) and systematic ratings of relevant personality traits, such as excitability. The Canine Behavioral Assessment & Research Questionnaire (CBARQ; Hsu and Serpell 2003) and the Dog Personality Questionnaire (Jones 2008) are two validated tools which have been created to assess longer-term individual differences in behavior and temperament in dogs and could be useful in future work. While the current study found evidence for group-level differences in performance, a more in-depth picture of each dog's temperamental profile could allow for predictions on an individual level. In past work investigating problem solving in dogs, Marshall-Pescini et al. (2008) found a significant correlation between dogs' successful performance and owner-reported temperament measures of high trainability and little to no stranger-directed fear. Even more to the point, Fox and Stelzner (1966) found that puppies who had not been handled, and thus were prone to emotional arousal, were worse at solving a detour task than their handled littermates. These emotionally aroused puppies were more likely to run into the barrier with their noses, similar to what we coded as an inhibitory failure in our own study. In their experiment, the puppies' arousal was related to temperamental differences that arose from controlled differences in the puppies' early rearing environment (Fox and Stelzner 1966).

In addition, future research could try to measure positive versus negative perception of the arousing stimulus in order to determine what effect that might have. In animal work, it can be difficult to determine the valence of the stimulus to an individual animal. For example, in the past, tail wagging has been used as a measure of both positive and negative arousal in dogs (Freedman et al. 1961; Pluijmakers et al. 2010; Rehn and Keeling 2011; Rehn 2013). However, recent studies suggest that the laterality of tail-wagging provides a window into the dog's emotional state, with a left-biased wag corresponding to positive, approach-worthy situations and a right-biased wag corresponding to threating, withdrawal-producing situations (Quaranta et al. 2007). These findings indicate that tail wagging might be a good candidate measure to answer the question of if and how the valence of arousal matters.

In conclusion, it appears that formal training and artificial selection can potentially lead to problem-solving biases that are moderated by temperament (i.e., Hare et al. 2005). These findings open the door for future research to further examine the role of learning and development in inhibitory control to help elucidate the circumstances in which animals can best exercise such control.

Acknowledgments

Thanks to R. Seyfarth, D. Cheney, J. Serpell, A. Duckworth, and two anonymous reviewers for valuable feedback on drafts and statistical analysis, as well as A. Gersick, I. Schamberg, and N. Snyder-Mackler for thoughtful discussion. We also thank K. Morucci and L. Lewis for help with coding, along with K. Duffy, E. Blumstein, T. Jones, A. Reinhardt, L. Thielke, L. Strassberg, M. Jackson, C. Wang, and members of the Duke Canine Cognition Center for assistance with data collection and testing. We're extremely grateful to the administration, trainers, and staff at Canine Companions for Independence and especially Paul Mundell, former National Director of Canine Programs and current Chief Executive Officer, for allowing us to work with their dogs and use their state-of-the-art facilities. This work was supported in part by the Vertical Integration Program, the Duke Undergraduate Research Support Office, the AKC Canine Health Foundation, a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1321851, Office of Naval Research Grant No. NOOO14-12-1-0095, and National Institute of Health Grant 5 R03 HD070649-02. Any opinions, findings, and conclusions or recommendations expressed in this material are solely the responsibility of the authors and do not necessarily reflect the views of the National Science Foundation or the AKC Canine Health Foundation.

Compliance with Ethical Standards

Funding: The study was funded by the AKC Canine Health Foundation, the National Science Foundation (Grant No. DGE-1321851), the Office of Naval Research (Grant No. NOOO14-12-1-0095), and the National Institute of Health (Grant No. 5 R03 HD070649-02). All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted.

References

Amici F, Aureli F, Call J (2008) Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. Curr Biol 18 (18):1415-1419

Anderson KJ (1990) Arousal and the inverted-u hypothesis: A critique of Neiss's "Reconceptualizing arousal". Psychol Bull 107 (1):96-100

Anderson KJ (1994) Impulsitivity, caffeine, and task difficulty: A within-subjects test of the Yerkes-Dodson law. Pers Indiv Differ 16 (6):813-829

Blair C, Diamond A (2008) Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. Dev Psychopathol 20 (3):899-911

Boogert NJ, Anderson RC, Peters S, Searcy WA, Nowicki S (2011) Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. Anim Behav 81 (6):1209-1216

Bray EE, MacLean EL, Hare BA (2014) Context specificity of inhibitory control in dogs. Anim Cogn 17 (1):15-31

Broadhurst P (1957) Emotionality and the Yerkes-Dodson law. J Exp Psychol 54 (5):345-352

Broadhurst P (1959) The interaction of task difficulty and motivation: The Yerkes-Dodson Law revived. Acta Psychol 16:321-338

Cieri RL, Churchill SE, Franciscus RG, Tan J, Hare B (2014) Craniofacial feminization, social tolerance, and the origins of behavioral modernity. Curr Anthropol 55 (4):419-443

Cole LW (1911) The relation of strength of stimulus to rate of learning in the chick. Journal of Animal Behavior 1 (2):111-124

Crawley MJ (2005) Statistics: An Introduction using R. John Wiley & Sons, Ltd., London

Diamond (2010) The evidence base for improving school outcomes by addressing the whole child and by addressing skills and attitudes, not just content. Early Educ Dev 21 (5):780-793

Diamond A (1990) Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of, Inhibitory Control in Reaching. Ann N Y Acad Sci 608 (1):637-676

Diamond A, Zola-Morgan S, Squire LR (1989) Successful performance by monkeys with lesions of the hippocampal formation on AB and object retrieval, two tasks that mark developmental changes in human infants. Behav Neurosci 103 (3):526

Dodson J (1915) The relation of strength of stimulus to rapidity of habit-formation in the kitten. Journal of Animal Behavior 5 (4):330-336

Dodson J (1917) Relative values of reward and punishment in habit formation. Psychobiology 1 (3):231-276

Duffy E (1957) The psychological significance of the concept of "arousal" or "activation". Psychol Rev 64 (5):265-275

Eysenck HJ (2002) The Dynamics of Anxiety & Hysteria: An Experimental Application of Modern Learning Theory to Psychiatry. Transaction Publishers, New Brunswick

Fox MW, Stelzner D (1966) Behavioural effects of differential early experience in the dog. Anim Behav 14 (2–3):273-281

Frank H, Frank MG (1982) Comparison of problem-solving performance in six-week-old wolves and dogs. Anim Behav 30 (1):95-98

Freedman DG, King JA, Elliot O (1961) Critical period in the social development of dogs. Science 133 (3457):1016-1017

Hare, Plyusnina I, Ignacio N, Schepina O, Stepika A, Wrangham R, Trut L (2005) Social cognitive evolution in captive foxes is a correlated by-product of experimental domestication. Curr Biol 15 (3):226-230

Hare, Tomasello M (2005a) Human-like social skills in dogs? Trends Cogn Sci 9 (9):439-444

Hare B, Melis AP, Woods V, Hastings S, Wrangham R (2007) Tolerance allows bonobos to outperform chimpanzees on a cooperative task. Curr Biol 17 (7):619-623

Hare B, Tomasello M (2005b) The emotional reactivity hypothesis and cognitive evolution: Reply to Miklósi and Topál. Trends Cogn Sci 9 (10):464-465

Hare B, Wobber V, Wrangham R (2012) The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. Anim Behav 83:573-585

Hebb DO (1955) Drives and the CNS (conceptual nervous system). Psychol Rev 62 (4):243-254

Hernádi A, Kis A, Turcsán B, Topál J (2012) Man’s underground best friend: domestic ferrets, unlike the wild forms, show evidence of dog-like social-cognitive skills. PLoS ONE 7 (8):e43267

Hsu Y, Serpell JA (2003) Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. J Am Vet Med Assoc 223 (9):1293-1300

Jones AC (2008) Development and validation of a Dog Personality Questionnaire. Unpublished doctoral dissertation, University of Texas at Austin, Austin

Kagan J, Snidman N (2004) The Long Shadow of Temperament. Belknap Press, USA

MacLean E, Hare B (2013) Spontaneous triadic engagement in bonobos (*Pan paniscus*) and chimpanzees (*Pan troglodytes*). J Comp Psychol 127 (3):245-255

MacLean EL, Hare B, Nunn CL, Addessi E, Amici F, Anderson RC, Aureli F, Baker JM, Bania AE, Barnard AM (2014) The evolution of self-control. P Natl Acad Sci USA 111 (20):E2140-E2148

Marshall-Pescini S, Valsecchi P, Petak I, Accorsi PA, Previde EP (2008) Does training make you smarter? The effects of training on dogs' performance (*Canis familiaris*) in a problem solving task. Behav Process 78 (3):449-454

Marshall-Pescini S, Virányi Z, Range F (2015) The Effect of Domestication on Inhibitory Control: Wolves and Dogs Compared. PloS one 10 (2):e0118469

McGowan RT, Rehn T, Norling Y, Keeling LJ (2014) Positive affect and learning: exploring the "Eureka Effect" in dogs. Anim Cogn 17 (3):577-587

McGreevy PD, Georgevsky D, Carrasco J, Valenzuela M, Duffy DL, Serpell JA (2013) Dog behavior co-varies with height, bodyweight and skull shape. PLoS ONE 8 (12):e80529

Melis AP, Hare B, Tomasello M (2006) Engineering cooperation in chimpanzees: tolerance constraints on cooperation. Anim Behav 72 (2):275-286

Miklósi Á (2007) Dog behaviour, evolution, and cognition. Oxford University Press, Oxford

Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington HL, Houts R, Poulton R, Roberts BW, Ross S (2011) A gradient of childhood self-control predicts health, wealth, and public safety. P Natl Acad Sci USA 108 (7):1-6

Osthaus B, Marlow D, Ducat P (2010) Minding the gap: spatial perseveration error in dogs. Anim Cogn 13 (6):881-885

Parker KJ, Buckmaster CL, Justus KR, Schatzberg AF, Lyons DM (2005) Mild early life stress enhances prefrontal-dependent response inhibition in monkeys. Biol Psychiat 57 (8):848-855

Parker KJ, Buckmaster CL, Lindley SE, Schatzberg AF, Lyons DM (2012) Hypothalamic-pituitary-adrenal axis physiology and cognitive control of behavior in stress inoculated monkeys. Int J Behav Dev 36 (1):45-52

Pluijmakers JJ, Appleby DL, Bradshaw JW (2010) Exposure to video images between 3 and 5 weeks of age decreases neophobia in domestic dogs. Appl Anim Behav Sci 126 (1):51-58

Pongrácz P, Miklósi Á, Kubinyi E, Gurobi K, Topál J, Csányi V (2001) Social learning in dogs: the effect of a human demonstrator on the performance of dogs in a detour task. Anim Behav 62 (6):1109-1117

Prescott MJ, Morton DB, Anderson D, Buckwell A, Heath S, Hubrecht R, Jennings M, Robb MD, Ruane MB, Swallow MJ (2004) Refining dog husbandry and care: Eighth report of the BVA AWF/FRAME/RSPCA/UFAW joint working group on refinement. Lab Anim 38:S1-S90

Quaranta A, Siniscalchi M, Vallortigara G (2007) Asymmetric tail-wagging responses by dogs to different emotive stimuli. Curr Biol 17 (6):R199-R201

Rehn T (2013) Best of Friends? Investigating the Dog-Human Relationship. Swedish University of Agricultural Sciences, Uppsala

Rehn T, Keeling LJ (2011) The effect of time left alone at home on dog welfare. Appl Anim Behav Sci 129 (2):129-135

Revelle W, Loftus DA (1992) The implications of arousal effects for the study of affect and memory. In: Christianson S-A (ed) The handbook of emotion and memory: Research and theory. Psychology Press, pp 113-149

Rosati AG, Hare B (2013) Chimpanzees and bonobos exhibit emotional responses to decision outcomes. PLoS ONE 8 (5):e63058

Schlosberg H (1954) Three dimensions of emotion. Psychol Rev 61 (2):81-88

Serpell JA, Duffy DL (2014) Chapter 2: Dog Breeds and Their Behavior. In: Horowitz A (ed) Domestic Dog Cognition and Behavior. Springer-Verlag, Berlin, pp 31-57

Tooby J, Cosmides L (2005) Conceptual foundations of evolutionary psychology. In: Buss DM (ed) The handbook of evolutionary psychology. John Wiley & Sons, Inc., Hoboken, pp 5-67

Topál, Erdőhegyi Á, Mányik R, Miklós Á (2006) Mindreading in a dog: an adaptation of a primate mental attribution study. Int J Psychol Psychol Ther 6 (3):365-379

Topál J, Miklósi Á, Gácsi M, Dóka A, Pongrácz P, Kubinyi E, Virányi Z, Csányi V (2009) Chapter 3 The Dog as a Model for Understanding Human Social Behavior. In: Brockmann JH, Roper TJ, Naguib M, Wynne-Edwards KE, Mitani JC, Leigh WS (eds) Advances in the Study of Behavior, vol 39. Academic Press, pp 71-116

van der Meere J, Stemerdink N, Gunning B (1995) Effects of presentation rate of stimuli on response inhibition in ADHD children with and without tics. Percept Motor Skill 81 (1):259-262

Vlamings PHJM, Hare B, Call J (2010) Reaching around barriers: the performance of the great apes and 3–5-year-old children. Anim Cogn 13 (2):273-285

Watters PA, Martin F, Schreter Z (1997) Caffeine and cognitive performance: The nonlinear Yerkes–Dodson law. Hum Psychopharm Clin 12 (3):249-257

Wright HF, Mills DS, Pollux PMJ (2011) Development and validation of a psychometric tool for assessing impulsivity in the domestic dog (*Canis familiaris)*. Int J Comp Psychol 24 (2):210-225

Wright HF, Mills DS, Pollux PMJ (2012) Behavioural and physiological correlates of impulsivity in the domestic dog (*Canis familiaris*). Physiol Behav 105 (3):676-682

Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 18 (5):459-482

Table 1

Descriptive statistics for pet dog participants (N = 30).

Dog Name	Breed	Order	Sex	Age (months)
Lucky	Mixed: Rottie/Cattle Dog	A	M	73.4
Dooright	Golden Retriever	A	M	80.3
Tanuk	Alaskan Malamute	A	M	36
Jaq	Rat Terrier	A	M	93.8
Scout	Beagle	A	F	115.8
Taylor	Pug	A	F	55.5
Oscar	Mixed: Labrador	A	M	52.6
Carolina	Great Pyrenees	A	F	58.5
Sarah	Mixed: Terrier/Cattle	A	F	88.6
Cassidy	Irish Setter	A	F	15.7
Sienna	Vizsla	A	F	137.1
Layla	Mixed: Hound/Shepherd	A	F	94.1
Bugsy	Mixed: Pointer/Dane	A	M	75.1
Autree	English Pointer	A	F	52.7
Blue	Mixed: Labrador/Chow	A	M	58.4
Merlin	Border Collie	B	M	30.3
Geisha	Mixed: Husky/Chow	B	F	57.3
Guga	Portuguese Water Dog	B	M	32.8
Enzo	Jack Russell Terrier	B	M	7.8
Disco	Mixed: Border Collie	B	M	67.5
Rogue	Mixed: Blue Tick	B	F	52.6
Max	Belgian Tervuren	B	M	96
Tola	Beagle	B	F	108
Bodie	Mixed: Collie/Chow	B	M	40.7
Loki	Chihuahua	B	M	54.2
Zeke	Border Collie	B	M	48
Lilah	Australian Shepherd	B	F	9.2
Deacon	Maltese	B	M	71
Lily	Poodle	B	F	74.5
Charlie Brown	Cavalier KC Spaniel	B	F	27

Table 2

Descriptive statistics for assistance dog participants (N =76).

Dog Name	Breed	Order	Sex	Age (months)
Gedeit	Labrador	A	M	26.8
Dex	Labrador-Golden Cross	A	M	27.2
Noland	Labrador-Golden Cross	A	M	23.9
Docker	Labrador-Golden Cross	A	M	24.8
Mulan	Labrador-Golden Cross	A	F	24
Greer	Labrador-Golden Cross	A	F	22.1
Dunbar	Labrador-Golden Cross	A	M	22.7
Lindsay	Labrador-Golden Cross	A	F	24.4
Brynna	Labrador-Golden Cross	A	F	23
Mardene	Labrador-Golden Cross	A	F	24.2
Tootsie	Labrador	A	F	25.6
Cabernet	Labrador	A	F	20.6
Oracle	Labrador	A	F	24.2
Zenrick	Labrador	A	M	27.9
Yazoo	Labrador-Golden Cross	A	M	30.6
Eva	Labrador-Golden Cross	A	F	25.2
Wendy	Labrador-Golden Cross	A	F	23.4
Heather	Labrador-Golden Cross	A	F	25
Safari	Labrador-Golden Cross	A	F	21.4
Hazel	Labrador-Golden Cross	A	F	27.4
Bramble	Labrador	A	F	25.8
Thelma	Labrador	A	F	23.8
Webb	Labrador-Golden Cross	A	M	23.6
Kaz	Labrador-Golden Cross	A	F	25.2
Katiya	Labrador-Golden Cross	A	F	22.8
Magnus	Labrador-Golden Cross	A	M	29.6
Bliss	Labrador-Golden Cross	A	F	23.7
Flavia	Labrador-Golden Cross	A	F	21
Fleur	Labrador-Golden Cross	A	F	21
Jetta	Labrador-Golden Cross	A	F	21
Claribel	Labrador-Golden Cross	A	F	30
Mojave	Labrador-Golden Cross	A	M	21.9

Burney	Labrador	A	M	20.6
Grove	Labrador	A	F	20.3
Neiman	Labrador	A	M	31.3
Daphne	Labrador-Golden Cross	A	F	20.6
Coraline	Labrador	A	F	20.8
Helen	Labrador-Golden Cross	A	F	20.5
Libby	Labrador-Golden Cross	A	F	22.5
Torelyn	Labrador	A	F	23.8
Minos	Labrador-Golden Cross	A	M	22.4
Lefty	Labrador-Golden Cross	A	M	22.5
Wonder	Labrador-Golden Cross	A	F	26.1
Winnie	Labrador-Golden Cross	A	F	21.4
Novi	Labrador-Golden Cross	A	F	22.4
Veronica	Labrador-Golden Cross	A	F	21.5
Freedom	Labrador-Golden Cross	B	M	20.4
Star	Labrador-Golden Cross	B	F	21.1
Wilde	Labrador-Golden Cross	B	M	20.7
Beula	Labrador	B	F	25.4
Oreo	Labrador	B	F	28.8
Fitz	Labrador-Golden Cross	B	M	25
Newkirk	Labrador-Golden Cross	B	M	24.1
Rodney	Labrador	B	M	26
Jovi	Labrador-Golden Cross	B	F	20.4
Kanga	Labrador-Golden Cross	B	F	20
Neffa	Labrador-Golden Cross	B	M	22
Judge	Labrador-Golden Cross	B	M	20.5
Gill	Labrador	B	M	20.6
Wayne	Labrador-Golden Cross	B	M	20.9
Halex	Labrador-Golden Cross	B	F	25.1
Chrissie	Labrador	B	F	23.2
Hydra	Labrador-Golden Cross	B	F	25.1
Nolan	Labrador-Golden Cross	B	M	22
Neptune	Labrador-Golden Cross	B	M	22.2
Lightning	Labrador-Golden Cross	B	M	22.6
Rayleigh	Labrador-Golden Cross	B	F	21.7
Vonne	Labrador-Golden Cross	B	F	21.3
Kelsey	Labrador-Golden Cross	B	F	22.6

Peter	Labrador	B	M	22
Wilfred	Labrador-Golden Cross	B	M	21.2
Kiri	Labrador-Golden Cross	B	F	22.7
Marina	Labrador-Golden Cross	B	F	22.5
Rapunzel	Labrador-Golden Cross	B	F	22
Stanford	Labrador	B	M	19.6
Nadia	Labrador-Golden Cross	B	F	22.5

Table 3

Results of a Linear Mixed Model in which the dependent variable was the composite score.

Predictor variables	Estimate	SE	*t* value	*p* value
Population	-0.32	0.12	-2.63	0.0095**
Order	-0.25	0.09	-2.91	0.0045**
Trial number	-0.09	0.01	-14.28	0.0000***
Trial type	-0.28	0.04	-6.35	0.0000***
Population x trial type	0.70	0.08	8.62	0.0000***
Population x order	0.47	0.16	2.96	0.0038**

Predictor variables were population (pet vs. assistance), order (low arousal trials first vs. low arousal trials first), trial number (1-10), and trial type (low arousal vs. high arousal). Dog ID was entered as a random effect. N = 30 pet dogs and 76 assistance dogs.

*** $p < 0.001$; ** $p < 0.01$

Fig. 1 The predictions of the Yerkes-Dodson hypothesis. For a simple task, it posits a positive linear relationship between arousal level and task performance. For a complex task, it posits an inverted U-shaped relationship, wherein increasing arousal level is linked to stronger performance only up to a certain point, after which increasing arousal harms performance

Fig. 2 **(A)** Curtain apparatus from the dog's perspective **(B)** The position of the experimenter while calling the dog behind the curtain apparatus during both low arousal and high arousal trials

Fig. 3 **(A)** Performance of assistance dogs on the detour arousal task by trial number and trial type. The lines represent the mean composite response score (touch + pathway + time to success), which is an inhibitory control failure index in which higher scores correspond to longer and less efficient problem solving. The gray line indicates dogs (n=46) who experienced order A, High Arousal First (5 high arousal detour trials followed by 5 low arousal detour trials), while the black line indicates dogs (n = 30) who experienced order B, Low Arousal First (5 low arousal detour trials followed by 5 high arousal detour trials); **(B)** Performance of pet dogs on the detour arousal task by trial number and trial type. The gray line indicates dogs (n=15) who experienced order A, High Arousal First, while the black line indicates dogs (n=15) who experienced order B, Low Arousal First

Fig. 4 **(A)** Cumulative performance of assistance (n=76) and pet (n=30) dogs during low arousal and high arousal trials. The bars represent the mean composite response score (touch + pathway + time to success), which is an inhibitory control failure index in which higher scores correspond to longer and less efficient problem solving. The interaction between trial type and dog type is significant ($p < 0.001$), with assistance dogs exhibiting optimal levels of inhibitory control during high arousal trials and pet dogs exhibiting optimal levels during low arousal trials; **(B)** Cumulative performance of assistance (n=76) and pet (n=30) dogs over the entire task, divided into groups based on those that completed high arousal trials first and those that completed low arousal trials first. The bars represent the same as in part (A). The interaction between order and dog type is significant, with assistance dogs that completed high arousal trials first exhibiting optimal levels of inhibitory control on the task overall and pet dogs that completed low arousal trials first exhibiting optimal levels of inhibitory control on the task overall ($p < 0.01$)

Fig. 5 Average performance of dogs in the 1st quintile of baseline arousal (n=23), 3rd quintile of baseline arousal (n=21), and 5th quintile of baseline arousal (n=22). Baseline arousal was determined by tail-wagging rate during the familiarization walk-around, prior to the start of testing. Performance is shown as a difference score, acquired by taking the mean composite response score (touch + pathway + time to success) for high arousal trials and subtracting the mean composite response score for low arousal trials. Negative scores correspond to better performance under high arousal conditions, scores close to zero correspond to no strong difference between conditions, and positive scores correspond to better performance under low arousal conditions.

Fig. 1

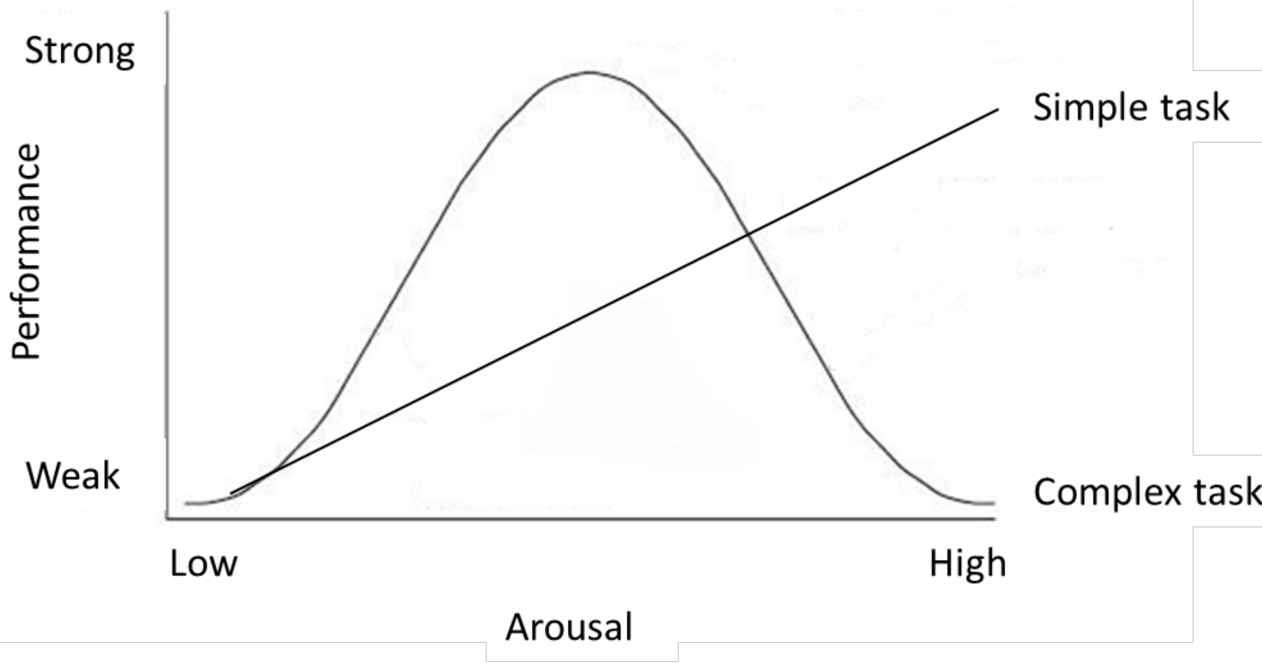
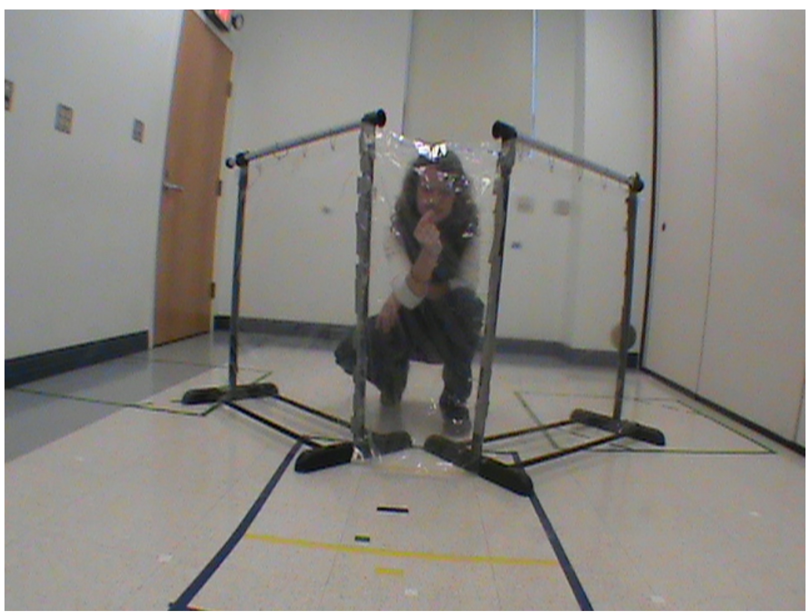

Strong
Performance
Weak
Simple task
Complex task
Low
High
Arousal

Fig. 2

a)



b)

Fig. 3

a) b)

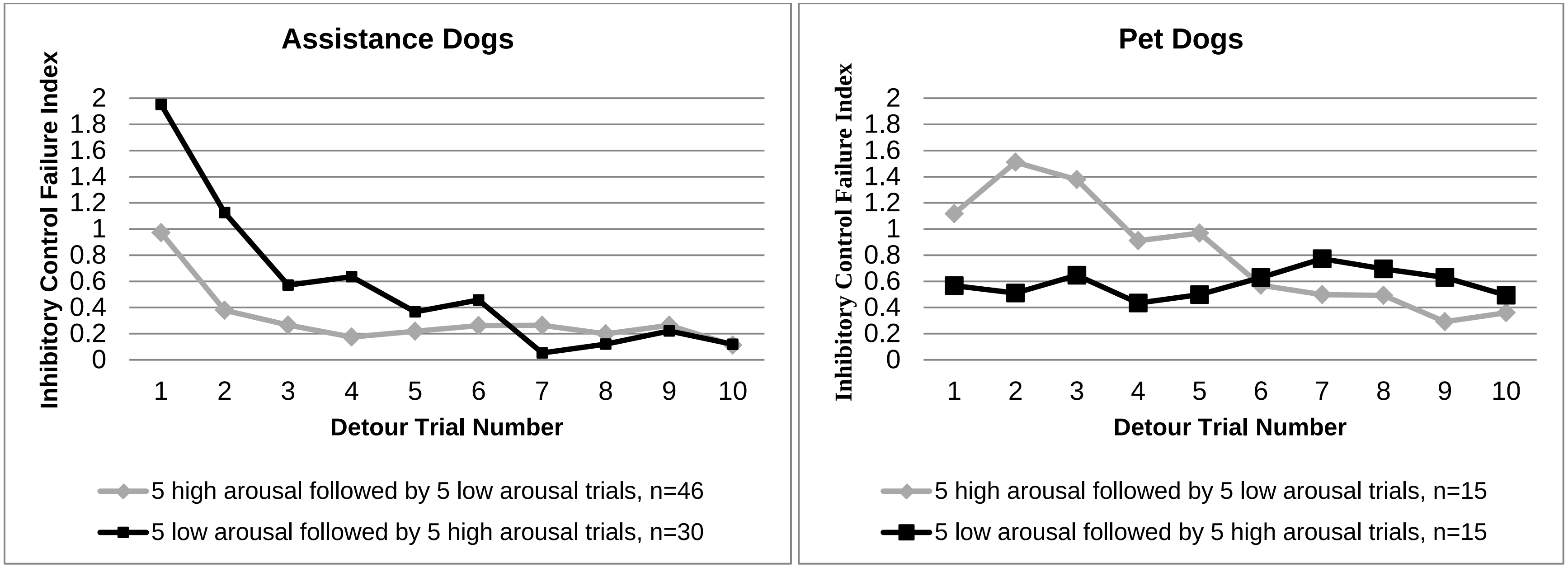
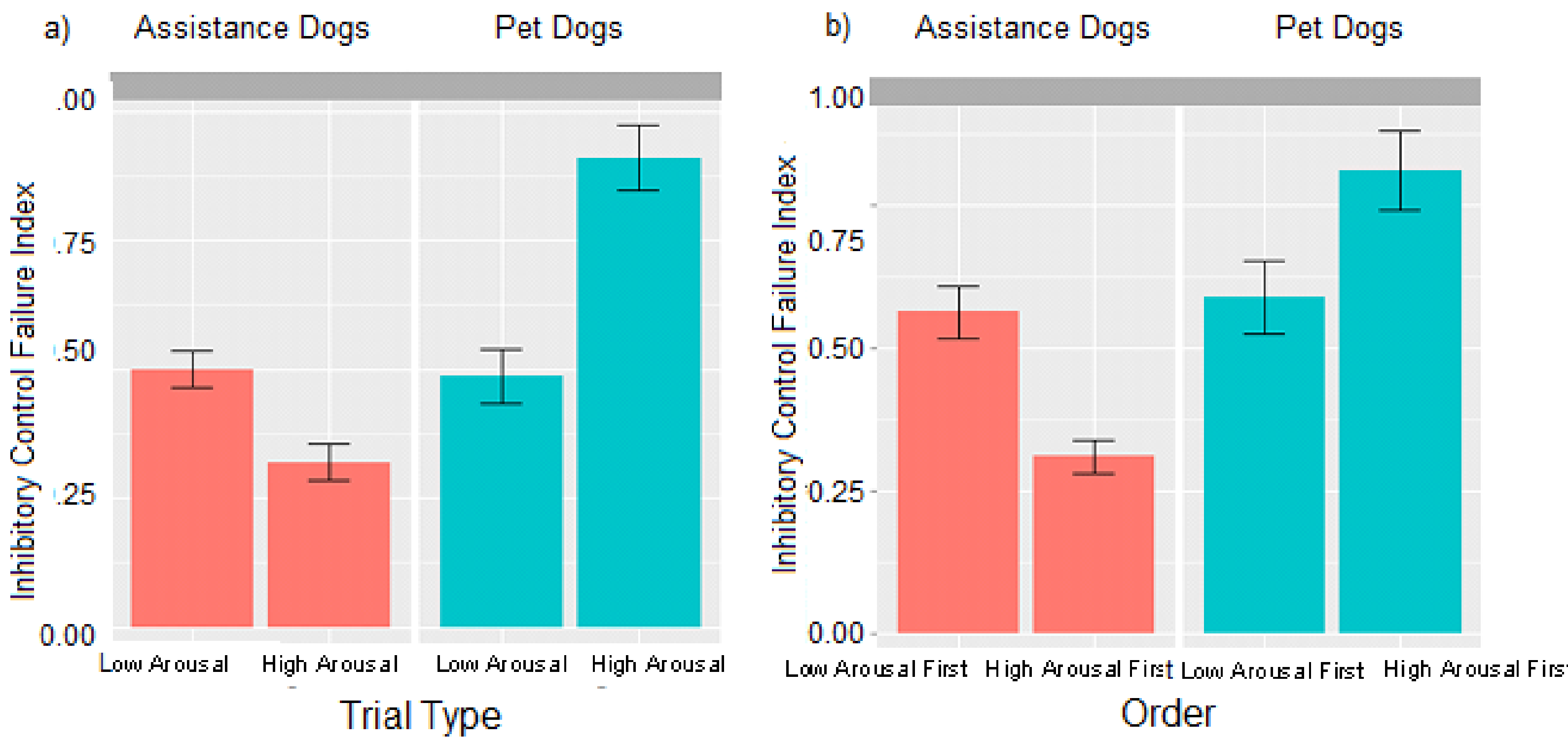
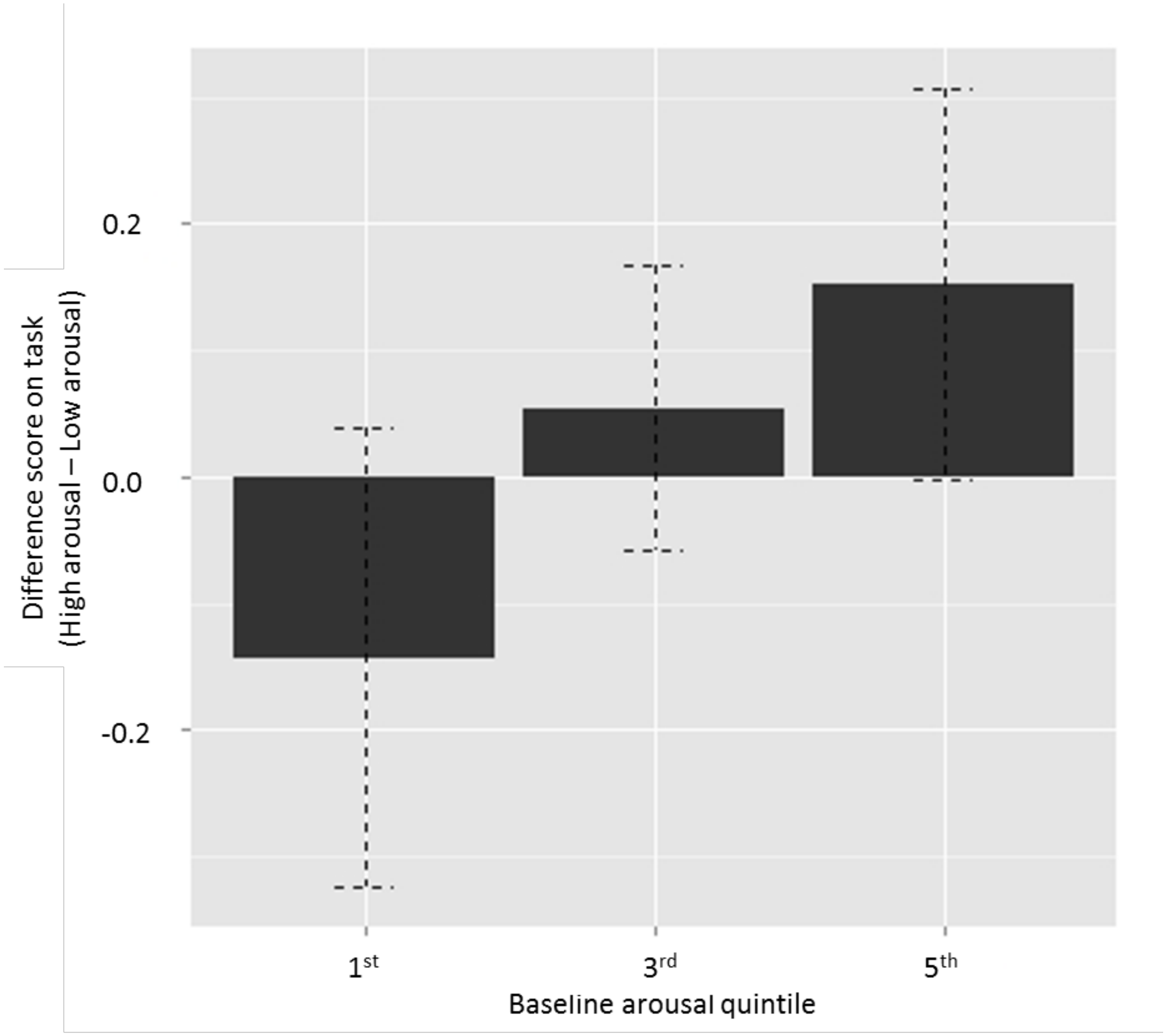

Assistance Dogs
Inhibitory Control Failure Index
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10
Detour Trial Number
5 high arousal followed by 5 low arousal trials, n=46
5 low arousal followed by 5 high arousal trials, n=30
Pet Dogs
Inhibitory Control Failure Index
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10
Detour Trial Number
5 high arousal followed by 5 low arousal trials, n=15
5 low arousal followed by 5 high arousal trials, n=15

Fig. 4



a)
Assistance Dogs
Pet Dogs
Inhibitory Control Failure Index
.00
.75
.50
.25
0.00
Low Arousal
High Arousal
Low Arousal
High Arousal
Trial Type
b)
Assistance Dogs
Pet Dogs
Inhibitory Control Failure Index
1.00
0.75
0.50
0.25
0.00
Low Arousal First
High Arousal First
Low Arousal First
High Arousal First
Order

Fig. 5


Difference score on task
(High arousal – Low arousal)
0.2
0.0
-0.2
1st
3rd
5th
Baseline arousal quintile