

Do owners know how impulsive their dogs are?

Jeffrey R. Stevens¹, Madeline Mathias¹, Megan Herridge¹, Kylie Hughes-Duvall¹, London M. Wolff¹, & McKenna Yohe¹

¹ University of Nebraska-Lincoln

Impulsivity is an important behavioral trait in dogs that affects many aspects of their relationship with humans. But how well do owners know their dog's levels of impulsivity? Two studies have investigated how owner perceptions of their dog's impulsivity correlate with the distance traveled in a spatial impulsivity task requiring choices between smaller, closer vs. larger, more distant food treats (Brady et al., 2018; Mongillo et al., 2019). However, these studies have demonstrated mixed results. The current project aimed to replicate these studies by correlating owner responses to the Dog Impulsivity Assessment Survey (DIAS) and the dog's maximum distance traveled in a spatial impulsivity task. We found that neither the DIAS overall score nor its three subcomponent scores correlated with dogs' distance traveled. This result replicates Mongillo et al.'s lack of a relationship but does not replicate Brady et al.'s effect, questioning the generalizability of owner reports of dog impulsivity. The lack of replication could result from differences in methodology and sample populations, but it raises intriguing questions about possible differences in dog characteristics and owner knowledge of their dogs across cultures.

Keywords: canine science, dog cognition, impulsivity, replication

Word count: 8,708

Introduction

When walking their dogs, owners will often encounter things that the dogs want to eat, chase, or roll in but the owner wants to avoid. To bypass those objects, dogs must control their impulsivity. Impulsivity is a multifaceted concept that typically captures an inability to wait, a preference for risky outcomes, a tendency to act without forethought, an insensitivity to consequences, and/or an inability to inhibit inappropriate behaviors (Evenden, 1999; Reynolds et al., 2006; Stevens, 2017b). While impulsivity can be evolutionarily adaptive in some circumstances (Stevens & Stephens, 2010; Fawcett et al., 2012), it is typically regarded as undesirable. Impulsivity is a critical component of dog behavior and cognition, with


implications for the dog-human relationship, including training, obedience, welfare, breeding, and re-homing. Therefore, considerable effort has been devoted to assessing owner perceptions of impulsivity (Vas et al., 2007; Wright et al., 2011), measuring impulsivity behaviorally (Wright et al., 2012; Bray et al., 2014; Riemer et al., 2014; Fagnani et al., 2016; Brucks, Marshall-Pescini, et al., 2017; van Horik et al., 2018), and determining biological markers of impulsivity (Hejjas et al., 2007; Kubinyi et al., 2012; Wright et al., 2012; Wan et al., 2013).

To measure impulsivity, Wright et al. (2011) developed the Dog Impulsivity Assessment Scale (DIAS), an owner-reported assessment of dog impulsivity with three components: behavioral regulation, aggression, and responsiveness. Owner ratings of dog impulsivity correlate with the dogs' choices in a temporal impulsivity task (Wright et al., 2012) and even correlate with performance on the same task six years later (Riemer et al., 2014). Thus, the DIAS seems to capture aspects of actual dog impulsive behavior. However, the temporal impulsivity task takes a long time to complete, so it cannot offer a quick measure of impulsivity. While the DIAS score does offer a quick measure (and it correlates with impulsive behavior), it can only be applied to dogs that have had an owner for enough time to assess impulsivity. Unfortunately, this assessment is difficult to use in shelter settings, where staff do not have enough interactions with a dog to assess impulsivity. Yet impulsivity is important to

Note: This is a pre-print and has not been peer reviewed.

PsyArXiv: <https://doi.org/10.31234/osf.io/hyv dq>

Version: 2022-05-20

Jeffrey R. Stevens,  <https://orcid.org/0000-0003-2375-1360>, Department of Psychology, Center for Brain, Biology & Behavior, University of Nebraska-Lincoln.

Correspondence concerning this article should be addressed to Jeffrey R. Stevens, B83 East Stadium, University of Nebraska, Lincoln, Lincoln, NE, USA 68588. ORCID 0000-0003-2375-1360. E-mail: jeffrey.r.stevens@gmail.com

assess in shelter situations to facilitate placing the dog in an appropriate home.

To develop a quick assessment of impulsivity, Brady et al. (2018) investigated how well DIAS scores matched a dogs' impulsive behavior in a *spatial impulsivity task* that involves choices between a smaller, closer reward and a larger, more distant reward (Stevens et al., 2005; Mühlhoff et al., 2011; Kralik & Sampson, 2012; Papale et al., 2012; Hopper et al., 2015). Spatial impulsivity tasks are quick to conduct and would offer a relatively easy measure of impulsivity. Brady et al. conducted three studies in which owners completed the DIAS and dogs completed a 'staircase' impulsivity task where they could choose between one and three food treats from two trays 25 cm away. When the dog chose the larger option, it was gradually moved farther away in the next trial until the dog switched to the smaller option. The researchers recorded the maximum distance traveled by the dog as a measure of impulsivity. Their Study 1 involved adult dogs in the lab, and they found that dogs whose owners reported low impulsivity traveled a farther distance in the spatial task. Their Study 2 replicated this design in a field setting and again found a correlation between DIAS score and maximum distance traveled by the dog. Their Study 3 tested 2–5 month-old puppies in the lab but did not find a correlation between owner-reported impulsivity and performance on the spatial impulsivity task.

Independently, Mongillo et al. (2019) tested a similar question by assessing owner-reported impulsivity with the DIAS and measuring dog impulsivity in a spatial task. The task, however, differed slightly from Brady et al. (2018). Instead of using a staircase method to find a maximum distance traveled, Mongillo et al. always had the larger option (seven treats) fixed at the farthest distance (3.6 m), and the smaller option (one treat) was randomly placed at one of four distances (1.8, 2.6, 3.0, and 3.6 m). They measured the proportion of choices for the larger reward at each distance. They, however, did not find a relationship between owner-reported impulsivity and performance on the spatial impulsivity task.

Given the mixed results between Brady et al. (2018) and Mongillo et al. (2019), the aim of our study was to directly replicate Brady et al.'s methods (Mongillo et al.'s paper was not published when our project began). Thus, in two studies, we assessed owner reports of dog impulsivity using the DIAS and had dogs complete a staircase-based spatial impulsivity task. We tested Brady et al.'s hypothesis that owner-reported impulsivity correlated with their dog's impulsivity in the spatial task.

Methods

To the best of our abilities, these methods replicated those of (Brady et al., 2018). Any specific deviations will be noted in the text.

Participants

For Study 1, we recruited participants through the Canine Cognition and Human Interaction Lab database from Nov 2018–Jul 2021. Recruitment was paused from Mar 2020–Apr 2021 due to the COVID-19 pandemic. We recruited 117 dog-owner pairs to the Canine Cognition and Human Interaction Lab facilities on the University of Nebraska-Lincoln campus, but 42 dogs did not begin the experiment due to lack of interest in treats or inability to engage in or pass training, leaving 75 dogs that started the experiment. Of those who started the experiment, 65 dogs completed the experiment and had complete survey data from the owner. For dog owners (participants), 49 identified as female, 15 as male, and 0 as non-binary (see Table S1 for more demographic information). Dogs (subjects) ranged in age from 0.5–16 years (mean \pm SD = 5.3 \pm 3.2) with 0 intact females, 30 spayed females, 3 intact males, and 32 neutered males.

For Study 2, we recruited participants through the Kenl Inn dog daycare client list and the Canine Cognition and Human Interaction Lab database from Aug 2020–Oct 2021. We recruited 103 dog-owner pairs to participate at the Kenl Inn facilities, but 54 dogs did not begin the experiment due to lack of interest in treats or inability to engage in or pass training, leaving 49 dogs that started the experiment. Of those, 43 dogs completed the experiment (15 from Kenl Inn daycare and 28 from our database). For dog owners, 39 identified as female, 4 as male, and 0 as non-binary (Table S1). Dogs ranged in age from 0.5–13 years (mean \pm SD = 3.6 \pm 3.0) with 2 intact females, 19 spayed females, 0 intact males, and 22 neutered males.

Ethics

All procedures were conducted in an ethical and responsible manner, in full compliance with all relevant codes of experimentation and legislation and were approved by the UNL Internal Review Board (protocols #17922 and 20491) and the Institutional Animal Care and Use Committee (protocol # 1703). All participants offered consent to participate, and they acknowledged that de-identified data could be published publicly.

Materials

For Study 1, the behavioral tasks occurred in a 5.1 x 6.5 m room with video cameras mounted in all four corners. The experimental set-up included a starting area for the dog, a curtain to occlude the dog's view during trial set-up, an opaque barrier running perpendicular to the curtain, and plates for food rewards (Figure 1a). We placed rewards on two equally sized circular plates of different colors (white or yellow) to

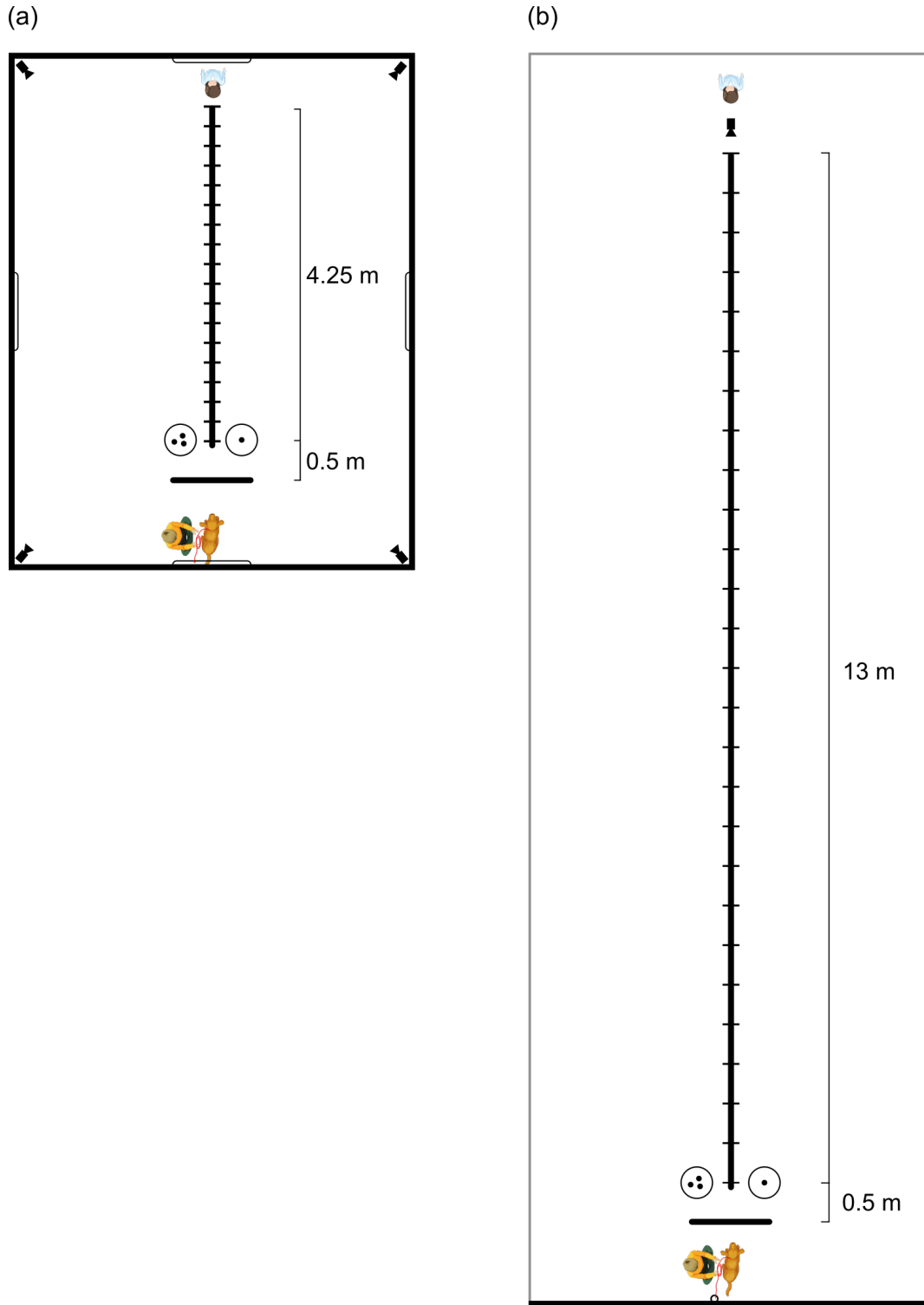


Figure 1. Experimental set-up. (a) For Study 1, the illustration shows the entire room size. (b) Study 2 was conducted in a much larger room, so only one wall (bottom) is illustrated. Dog figure from vecteezy.com and experimenter figures adapted from art by macrovector on freepik.com.

enhance discrimination of the reward amounts. On each plate, we placed either the larger reward consisting of three small (1 cm³) dog training treats or the smaller reward of one dog treat. The color of the plate containing each reward and the side of the small/large reward (right/left) were consistent within subjects across training and testing and were counterbalanced between subjects. Tape on the floor marked distances from the starting position in 0.25 m increments to a length of 4.25 m. Note that this distance is shorter than the maximum used by Brady et al. (2018) of 10 m. An opaque barrier running the length of the distances (4.25 m long, 50 cm high) ensured that once the dog selected an option, it could not see or obtain the unchosen option. An opaque curtain (80 cm high) separated the dog's starting area from the choice area and prevented them from observing the trial set-up. Two dogs could see over the curtain, which could have allowed the order of treat placement to affect their choices. The starting area included a metal bar attached to the wall with a 16 m leash attached to it.

For Study 2, the behavioral tasks occurred in a large gym where we used space approximately 10 x 16 m in size. The experimental set-up was very similar to that of Study 1 except (1) the distance ran up to 13 m (longer than Brady et al.'s (2018) length of 10 m), (2) the distances increased in steps of 0.5 m marked on a rope that ran the length of the 13 m center barrier, and (3) a single camera recorded the session from the end of the barrier (Figure 1b). The starting area included a metal hook attached to the wall with the 16 m leash.

Surveys

For Study 1, while their dogs were completing the behavioral task, owners completed a Qualtrics survey on a tablet computer in a waiting room outside of the experimental testing room (owners could observe the behavioral testing via an external monitor). The survey consisted of questions about dog and owner demographics, information about dog-owner separation, the dog's general behaviors, training, feeding, and exercise habits, and six scales. For the purposes of this study, we included the following measures: dog age, dog sex, dog neuter status, dog weight, dog AKC Canine Good Citizen status, dog training ratings, dog obedience ratings, dog problem behaviors, dog impulsivity, dog separation anxiety, dog-owner relationship ratings, owner personality, owner cognitive ability, owner gender, and owner household income. We calculated Revelle's omega total (ω_T) as our measure of internal consistency reliability of scales (Revelle & Zinbarg, 2008; McNeish, 2018) unless these computations failed (e.g., for scales with only two items), in which case, we calculate Cronbach's α (Table S2).

Dog Training. To assess dog training, we asked owners to rate how well trained they thought their dogs were on a scale from 1 to 10 with 10 being the best.

Dog Obedience and Behavioral Problems. We used the Bennett and Rohlf (2007) scale, which assessed dog behavior problems with 24 questions on a seven-point scale with five subscales: disobedience, aggression, nervousness, destructiveness, and excitability.

We also used Hiby et al.'s (2004) scale to assess obedience and problem behaviors in dogs. Obedience was assessed on a five-point scale with seven specific tasks and an overall obedience score. Behavioral problems were assessed by participants indicating whether their dog had never, previously, or currently shown 13 behavioral problems.

Dog Impulsivity. The Dog Impulsivity Assessment Scale (Wright et al., 2011) assessed impulsivity in dogs using a five-point scale (plus "don't know/not applicable"). The scale included 18 questions divided over three subscales (two questions are used in more than one subscale): behavioral regulation (Factor 1 in Wright et al. 2011), aggression (Factor 2), and responsiveness (Factor 3). We calculated the mean of the three subscales to generate a DIAS overall score (referred to as Overall Questionnaire Score or OQS in Wright et al. 2011).

Dog-Owner Relationship. The Monash Dog Owner Relationship Scale (Dwyer et al., 2006) used a seven-point scale to assess human-dog relationships by measuring how frequently owners engage in nine activities with their dogs.

Owner Personality. The brief Big-Five personality scale (Gosling et al., 2003) used a five-point scale to assess owner personality. The scale included 10 questions divided over five subscales: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. While some of the internal consistency reliability values were low, (1) there were only two items per subscale (which forced us to calculate Cronbach's α for reliability, as we could not compute Revelle's ω_T), (2) our values are similar to the original study, and (3) the test-retest reliability and convergent correlations with a ten-item inventory were quite high in the original study.

Owner Cognitive Ability. The Cognitive Reflection Task (Frederick, 2005) used three multiple-choice questions to assess cognitive reflection in owners. The Berlin Numeracy Test (Cokely et al., 2012) used four multiple choice questions to assess owner numeracy. Scores for both tests were calculated by summing the number of correct responses. We summed the scores from these two tests to generate an index of cognitive ability that ranged from 0-7 (Stevens et al., 2021). If participants failed to answer any of the seven questions, their cognitive ability score was excluded from analyses that included this score ($N = 26$).

For Study 2, owners completed a Qualtrics survey at home prior to bringing their dogs in for testing. The survey consisted of the following measures: dog age, dog sex, dog neuter status, dog weight, dog AKC Canine Good Citizen status, dog

training ratings, dog obedience ratings, dog impulsivity, dog separation anxiety, owner personality, owner gender, owner household income. We omitted the following measures from Study 1: dog problem behaviors, dog-owner relationship ratings, and owner cognitive ability.

Scales carried over from Study 1 included the Hiby et al. (2004) obedience score, the Dog Impulsivity Assessment Scale (Wright et al., 2011), and the brief Big-Five personality scale (Gosling et al., 2003). We replaced the training rating question with the C-BARQ trainability scale (Hsu & Serpell, 2003), which is an eight-item questionnaire on a five-point scale. Reliability values for scales are provided in Table S2.

Procedure

The experiment was conducted in a single approximately one-hour-long session. Dogs experienced a habituation phase, training phase, and testing phase. For Study 1, owners were with dogs during habituation but were outside the testing area during the training and testing phases. For Study 2, dogs were either dropped off for daycare or brought in for the study. If in daycare, the daycare staff brought the dogs over from the main building for testing and returned them after testing. For other dogs, the owner remained in their car while our staff retrieved the dog and brought them into the testing area. Therefore, owners were not present for testing in either study.

Habituation. To acclimate the dog to the space and equipment, one experimenter allowed the dog to walk around the waiting and testing areas for about 5 minutes. Dogs were allowed to sniff and walk around the curtain and barrier. Once dogs seemed accustomed to the space, two experimenters took the dog into the testing area. In Study 1, the owner left the testing area to take the survey in the waiting area, and the door between the two areas was closed. The experimenters then gave the dog treats to assess whether the dog was engaged with the experimenter or showed signs of stress from being separated from their owner. In Study 2, the habituation worked similarly to Study 1 but was slightly quicker because the owners were not present.

Training. The training phase occurred adjacent to the testing area to further acclimate the dog without the formal requirements of the testing situation (no curtain, no specific distances, etc.). In training, subjects chose between the smaller and larger reward amounts on plates. One experimenter (here labeled ‘presenter’) placed the rewards on the plates and presented them about 0.5 m away from the subject. The other experimenter (here labeled ‘handler’) sat next to the subject during the training, released the subject after presenting rewards, and encouraged them to choose a reward. Subjects were allowed to choose and consume one of the two options. The presenter immediately removed the unchosen option once the subject touched the chosen option. Training

session bouts consisted of 12 trials for Study 1 (matching Brady et al. (2018)) or 8 trials for Study 2 (shorter than Brady et al. to speed training and reduce treat consumption before the main experiment). For the subject to move on to the testing phase, they had to select the larger treat amount in at least 10 out of the 12 trials (Study 1) or 5 of 8 trials (Study 2). For Study 1, they also had to select the larger reward in all of the last five trials of one training bout. As soon as subjects completed the requisite number of trials (10 or 5), the bout was stopped. Subjects could attempt up to five (Study 1) or three (Study 2) training bouts to reach the passing criteria. If the subject failed these training criteria in the required number of training bouts, the test was concluded and the dog did not advance to the testing phase.

Testing. Testing began following a short break after meeting training criterion. For this phase, the handler brought the subject over to the starting area, attached a long leash to the dog’s collar, and directed the subject to face the curtain. Meanwhile, the presenter placed two plates 0.5 m from the starting position of the subject—one on either side of the central barrier. The same color of plates from the training phase were used here. The presenter then removed the curtain, walked to the opposite end of the barrier, and stood centered with the barrier with their back to the dog (Figure 1). The presenter alternated which side of the barrier they walked down each round of testing to reduce side bias in results. After 4 seconds to view both options, the handler gave the command that the owner typically used to release from a stay position or said “okay” or “go” and released the slack on the leash. The dog had 30 seconds to choose a plate after being released. If the dog refused to look at the plates, refused to move, or moved in a different direction from the plates during this time, the trial was marked as no choice. If the dog looked at the plates during the 30 seconds, the handler repeated the release command once more and the dog was given 60 more seconds to make a choice. A choice was defined as the plate that the subject first touched or ate from. After the dog consumed the treats on the chosen plate, the handler called the subject back to the starting position or led them back with the leash. The handler kept the dog from getting the unchosen rewards by holding onto the leash and guiding the dog back to the starting position. The presenter then replaced the curtain and replenished the treats on the selected plate. A video of testing trials is available at the Open Science Framework (<https://osf.io/eb5m3/>).

Each time the dog selected the larger reward option, the plate was moved back one step (0.25 m for Study 1 and 0.5 for Study 2) in the subsequent trial. The testing phase concluded once the dog selected the smaller reward five times in a row or refused to select an option either three times in a row or five times total throughout the duration of testing. The response variable for this measure was the maximum distance traveled for the large reward.

Separation anxiety. If the subject appeared anxious with signs such as panting, drooling, whining, cowering, shaking, or walking to the door during any of the phases, training and testing were paused to re-acclimate the subject to the environment. In Study 1, the door to the waiting area was opened to allow access to the waiting room where the owner stayed for 10 minutes. Owners were prompted to not interact with subject during this period, and rewards were only given to the subject by experimenters in the testing room. After 10 minutes, the owners were asked to stand in the testing room while still ignoring the subject. Then the door to the testing room was closed. For 5 minutes, the experimenters continued to give rewards to the subject while allowing the subject to explore the room off leash. After 5 minutes, the subject was put on the leash used in testing for a minute at a time alternating between being on and off leash. The owner was then asked to leave the testing room. With the door closed, experimenters continued to give rewards to the subject for 15 minutes with 1 minute intervals of the subject being on the leash dispersed throughout the 15-minute period. Subjects were then allowed to go home without continuing to the training or testing phase. If the dog appeared to be more comfortable taking treats in the testing room and being separated from their owner at the end of the extended habituation, the subject was invited back to restart the study again from the beginning on a different day. If the dog did not show improvement in its comfort level, the subject was not invited back. For Study 1, four dogs were invited for a second attempt, but only one completed testing. For Study 2, two dogs were invited back, and one completed testing.

Data analysis

We used R (Version 4.2.0; R Core Team, 2022) and the R-packages *BayesFactor* (Version 0.9.12.4.3; Morey & Rouder, 2018), *conflicted* (Version 1.1.0; Wickham, 2021), *gg-beeswarm* (Version 0.6.0; Clarke & Sherrill-Mix, 2017), *ggdist* (Version 3.1.1; Kay, 2022), *here* (Version 1.0.1; Müller, 2020), *metaBMA* (Version 0.6.7; Heck et al., 2019), *papaja* (Version 0.1.0.9999; Aust & Barth, 2020), *patchwork* (Version 1.1.1; Pedersen, 2020), *psych* (Version 2.2.3; Revelle, 2020), and *tidyverse* (Version 1.3.1; Wickham et al., 2019) for all our analyses. The manuscript was created using *rmarkdown* (Version 2.14, Xie et al., 2018) and *papaja* (Version 0.1.0.9999, Aust & Barth, 2020). Data, analysis scripts, supplementary materials, and the reproducible research materials are available at the Open Science Framework (<https://osf.io/eb5m3/>).

We calculated Bayes factors (BF_{10}) to provide the weight of evidence for the alternative hypothesis relative to the null hypothesis (Wagenmakers, 2007). For example, $BF_{10} = 3$ means that the evidence for the alternative hypothesis is 3 times stronger than the evidence for the null hypothesis. $BF_{10} = 1/3$ means that the evidence for the

null hypothesis is 3 times stronger than the evidence for the alternative hypothesis. Bayes factors between $1/3 - 3$ provide only anecdotal evidence, those between $3 - 10$ or $1/10 - 1/3$ provide moderate evidence, those between $10 - 100$ or $1/100 - 1/10$ provide strong evidence, and those above 100 or below $1/100$ provide very strong evidence (Andrzejewicz et al., 2015). Bayes factors were calculated with the `correlationBF()`, `ttestBF()`, and `anovaBF()` functions from the *BayesFactor* package using default (noninformative) priors.

The sample size for both studies was determined by an optional stopping rule based on Bayes factors (Schönbrodt et al., 2017). We calculated the Bayes factors for the correlation between maximum distance traveled and the DIAS overall score and stopped analysis when it reached either $1/3$ or 3 . We analyzed the first 20 subjects and checked our Bayes factor every 2-5 subjects to determine whether one of the thresholds was met. This resulted in our final sample size of 65 subjects for Study 1 and 43 subjects for Study 2.

Results

Impulsivity

Dogs traveled a mean \pm SD distance of $1.69 \text{ m} \pm 1.10 \text{ m}$ (range: $0.25 - 4.25 \text{ m}$) in Study 1 and $1.81 \text{ m} \pm 0.81 \text{ m}$ (range: $0.50 - 4.50 \text{ m}$) in Study 2 (Figure 2a). The mean \pm SD Dog Impulsivity Assessment Scale (DIAS) overall scores were 0.48 ± 0.09 (range: $0.31 - 0.70$) in Study 1 and 0.47 ± 0.06 (range: $0.36 - 0.63$) in Study 2 (Figure 2b).

Our primary analysis correlated the maximum distance traveled by the dog with the owner's DIAS overall score. Neither study showed evidence for a correlation between distance traveled and owner assessments of impulsivity (Study 1: $r = -0.10$, 95% CI $[-0.34, 0.15]$, $t(63) = -0.80$, $p = 0.429$, $BF_{10} = 0.38$; Study 2: $r = 0.04$, 95% CI $[-0.26, 0.34]$, $t(41) = 0.26$, $p = 0.798$, $BF_{10} = 0.35$; Figure 3).

For both studies, we collected data until the Bayes factor for this correlation reached either $1/3$ or 3 . Unfortunately, due to a coding error in the analysis script in which we did not remove sessions for training experimenters, we stopped both studies slightly before reaching either threshold (Study 1: $N = 65$, $BF_{10} = 0.38$; Study 2: $N = 43$, $BF_{10} = 0.35$). Because optional stopping is not a problem for Bayes factor analyses (Rouder, 2014), we can combine the data sets from the two studies to calculate Bayes factors for exploratory analyses of the overall effect. Despite the longer room and maximum distance in Study 2, the studies did not differ in either the distribution of distances traveled by the dogs ($\Delta M = -0.12$, 95% CI $[-0.49, 0.24]$, $t(104.68) = -0.66$, $p = 0.509$, $BF_{10} = 0.25$, Figure 2a) or the distribution of DIAS overall scores ($\Delta M = 0.01$, 95% CI $[-0.02, 0.04]$,

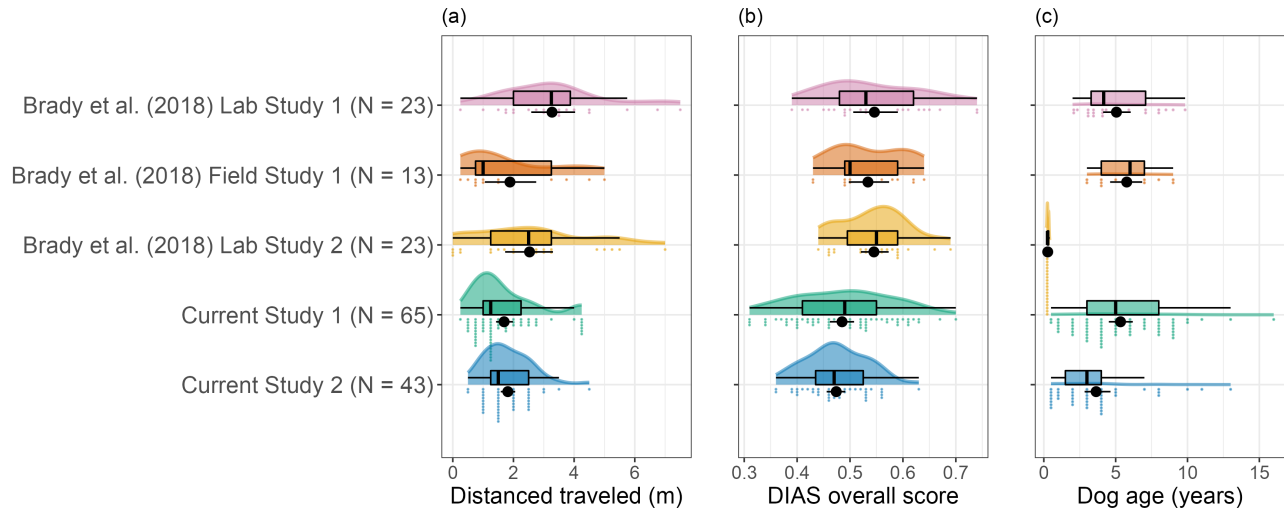


Figure 2. Distance traveled, DIAS overall score, and dog age across studies. Individual data points and summary statistics are shown for (a) distance traveled, (b) DIAS overall score, and (c) dog age for the Brady et al. (2018) and current studies. Colored dots represent individual dog data points, filled shapes represent density distributions, black dots and error bars represent mean and 95% confidence intervals, boxes represent interquartile range, lines within boxes represent medians, and whiskers represent 1.5 times the interquartile range.

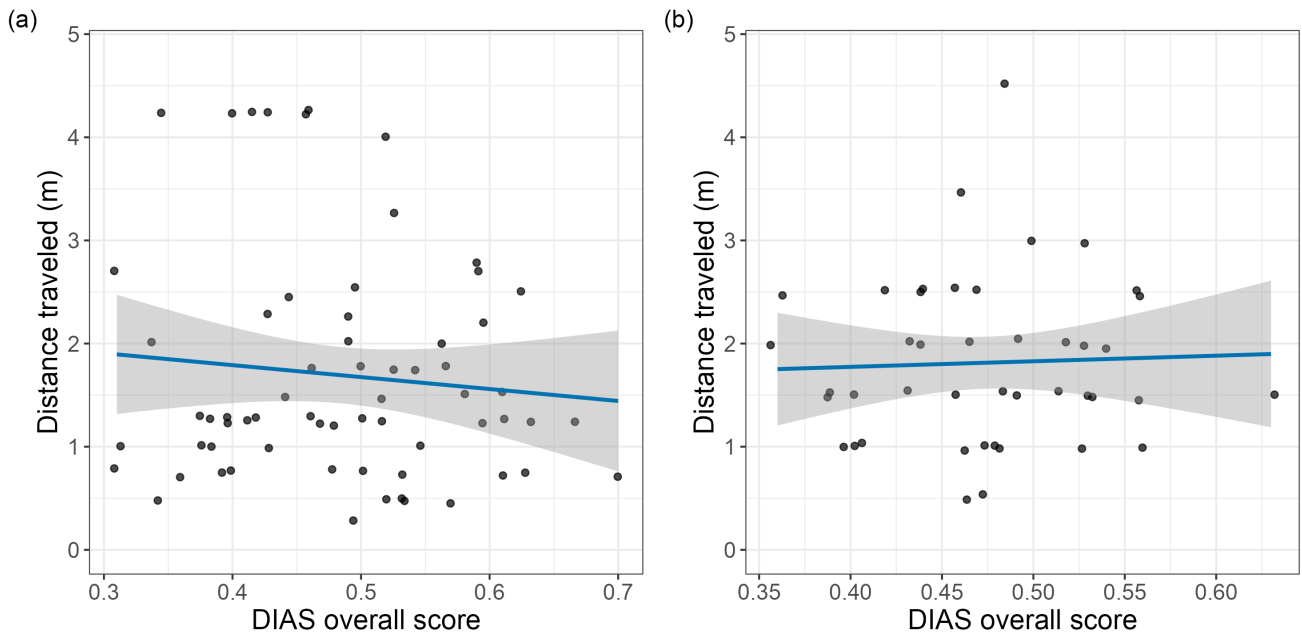


Figure 3. Relationship between distance traveled and DIAS overall score. Scatterplots of distance traveled and DIAS overall score for (a) Study 1 and (b) Study 2 showed no evidence for a correlation. Dots represent individual dog data points, lines represent best fitting linear regression models, and bands represent 95% confidence intervals around the regression models.

$t(105.85) = 0.74$, $p = 0.462$, $BF_{10} = 0.25$, Figure 2b). However, dogs in the second study were younger on average than those in the first study ($\Delta M = 1.70$, 95% CI [0.50, 2.90], $t(95.00) = 2.80$, $p = 0.006$, $BF_{10} = 5.77$, Figure 2c). With the combined data, there is moderate evidence for the null hypothesis of no correlation between dog travel distance and owner reports of impulsivity ($r = -0.07$, $BF_{10} = 0.28$).

Analyses of the DIAS subscales in both studies also showed (anecdotal) evidence for no correlations with distance traveled ($0.34 \leq BF_{10} \leq 0.46$; Figure S1). Combining the studies showed moderate evidence for no correlation between distance traveled and behavioral regulation ($r = -0.05$, $BF_{10} = 0.25$), aggression ($r = 0.07$, $BF_{10} = 0.28$), or responsiveness ($r = -0.07$, $BF_{10} = 0.28$).

Other measures

For exploratory analyses of other factors that may relate to distance traveled, we combined data from the two studies and only present Bayes factors for analyses. In terms of dog demographic characteristics, distance traveled did not relate to dog sex, weight, age, or AKC Canine Good Citizen status (Figure S2). For dog behavior measures, distance traveled did not relate to Bennett and Rohlf (2007) problem behavior scores, Hiby et al. (2004) obedience score, training score, or the owner report of separation anxiety (Figure S3). For owner characteristics, distance traveled did not relate to owner extraversion, agreeableness, stability, openness, cognitive ability, or whether they had other dogs in the household (Figure S4). We had insufficient evidence to robustly assess potential relationships with the Monash dog owner relationship scores and owner conscientiousness.

Study comparison

Because our findings did not replicate those of Brady et al. (2018), we conducted an exploratory comparison of their and our studies. Since our two studies do not differ in distance traveled or DIAS overall score (see Impulsivity section), we combined the data across studies for each of these measures. Our combined data have shorter distances traveled than Brady et al.'s Lab Study 1 ($\Delta M = -1.53$, 95% CI [-2.30, -0.76], $BF_{10} > 1000$) and their Lab Study 2 with puppies ($\Delta M = -0.79$, 95% CI [-1.65, 0.07], $BF_{10} = 7.49$) but comparable travel distances to their Field Study ($\Delta M = -0.14$, 95% CI [-1.15, 0.86], $BF_{10} = 0.32$; Figure 2a). Our DIAS overall score values were lower than those of their Lab Study 1 ($\Delta M = -0.07$, 95% CI [-0.11, -0.02], $BF_{10} = 26.96$) and their Lab Study 2 ($\Delta M = -0.06$, 95% CI [-0.10, -0.03], $BF_{10} = 47.65$), with anecdotal evidence for a difference with their Field Study ($\Delta M = -0.05$, 95% CI [-0.10, -0.01], $BF_{10} = 2.19$; Figure 2b).

To further compare studies, we conducted an exploratory meta-analysis of Brady et al.'s (2018) three studies, Mongillo et al.'s (2019) study, and our two studies. For Brady et al.'s studies and our studies, we used the provided correlation coefficients and sample sizes. For Mongillo et al., we used the correlation between DIAS scores and overall proportion choice for larger, more distant option. We converted these correlation coefficients to Fisher's z scale and calculated standard errors based on the equation $SE = 1/\sqrt{(N-3)}$ (Borenstein et al., 2009). We then input these values into the `meta_bma()` function in the *metaBMA* package (Heck et al., 2019) with default priors for Fisher's z values (Cauchy distribution with $scale=0.354$ for the effect size and inverse gamma with $shape=1$ and $scale=0.075$ for τ). This function allows us to test not only if there is evidence of an effect but also evidence for between-study heterogeneity (Gronau et al., 2021), that is, whether there is variation in true effect sizes across studies. We then converted the Fisher's z values back to correlation coefficients for presentation. This analysis used model averaging to calculate (1) a Bayesian estimate of the effect size across all studies, (2) a Bayes factor for evidence supporting the hypothesis of an effect, and (3) a Bayes factor for evidence supporting the presence of between-study heterogeneity (Gronau et al., 2021). This analysis found a model-averaged effect of $r = -0.11$ (95% credibility interval [-0.27, 0.04]), which provided anecdotal evidence of no correlation between distance traveled and DIAS score across all studies ($BF_{10} = 0.51$; Figure 4). It also found anecdotal evidence of no between-group heterogeneity ($BF_{10} = 0.74$).

Discussion

In two studies, we found no relationship between the distance traveled by dogs in a spatial impulsivity task and the owner reports of dog impulsivity. In fact, the distance traveled did not relate to any measures of dog behavior, dog characteristics, or owner characteristics.

Replication differences

Our two studies differed in four major ways. First, the training criteria differed, with subjects needing to choose the larger option in 10 of 12 trials in Study 1 and 5 of 8 trials in Study 2. For Study 2, choosing larger on 5 of 8 trials does not pass a binomial test (using $\alpha = 0.05$), so it is possible that subjects did not discriminate the amounts as well in Study 2. To examine whether the criteria difference influenced performance, we assessed whether subjects completed training faster in Study 2. Subjects required a mean \pm standard deviation of 1.58 ± 1.09 training bouts to advance in Study 1 and 1.31 ± 0.58 bouts in Study 2. Thus, there was anecdotal evidence for no difference in number of training bouts required across the two studies ($t(91.58) = 1.58$, $p = .117$, $BF_{10} = 0.51$). This difference

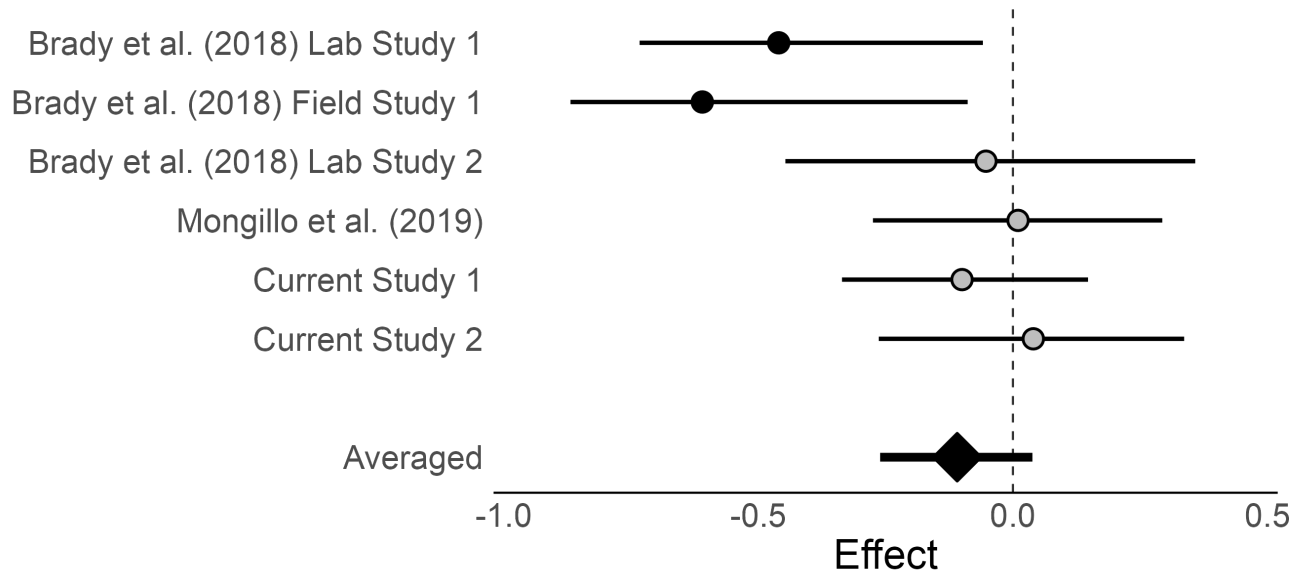


Figure 4. Forest plot of meta-analysis. Plot shows effect sizes (correlation coefficients, r) and 95% confidence/credibility intervals around effect sizes for Brady et al. (2018), Mongillo et al. (2019), and current studies. Black dots represent effect sizes different from 0, grey dots represent effect sizes not different from 0, and the black diamond represents the Bayesian model-averaged effect size over all studies. Study error bars represent confidence intervals, and model-averaged error bars represent credibility interval.

in training criteria likely did not influence the testing results since the distance traveled did not differ between studies. As a second difference, Study 2 used a longer maximum distance (13 m) and steps (0.5 m) in a larger room than Study 1 (maximum: 4.25 m, steps: 0.25 m). Study 1 used a distance more similar to Mongillo et al.'s (2019) maximum distance of 3.5 m, whereas Study 2 was more similar to Brady et al.'s (2018) maximum distance of 10 m. Despite this large difference in maximum distances and room size, the distance traveled did not differ between studies. A third difference between studies was that Study 1 drew participants exclusively from our database, whereas the second study included database participants and dog owners who used the dog daycare where we conducted the experiment. Thus, the daycare dogs were more familiar with the personnel and testing area than the other dogs. Nevertheless, we found anecdotal evidence that daycare and database dogs did not differ in the distance traveled or DIAS scores ($BFs < 1$), though the sample size for daycare dogs was small ($N = 15$). And, again, the distance traveled by dogs did not differ between studies. Finally, dogs in Study 2 were on average younger than those in Study 1. Yet, age did not influence distance traveled. Thus, though there were a few key differences between our studies, we reliably replicated the same lack of an effect across both studies.

We failed to replicate Brady et al.'s (2018) finding that owner reports of impulsivity matched their dogs' distance traveled in a spatial impulsivity task. Our results align with Mongillo et al.'s (2019) demonstration of no relationship, though we use

different methods to assess spatial impulsivity. Interpreting the inability to replicate results is difficult (Stanley & Spence, 2014; Maxwell et al., 2015), so we evaluate our replication based on LeBel et al.'s (2019) recommendations for replications. Our *methodological similarity* is high, as we directly aimed to replicate the methods of Brady et al. (2018). We used the same operationalizations of dependent and independent variables (DIAS overall score and maximum distance traveled), in contrast to Mongillo et al. (2019) who used different methods for eliciting choice and used proportion choice for the most distant option as the dependent variable. In terms of *replication differences*, we addressed a possible difference of a shorter distance in Study 1 by employing a longer distance in Study 2. Nevertheless, the distances traveled by dogs did not differ between our studies, though they were slightly shorter distances than Brady et al. found. Our DIAS overall scores appeared to be the same as Brady et al. There is no *investigator independence* issue, as all investigators from this study are completely independent from the original study. We aim for *study transparency* by posting all data on Open Science Framework (<https://osf.io/eb5m3/>); however, the study was not pre-registered. We facilitate *analytic result reproducibility* by posting our analysis scripts with our data, allowing others to reproduce our analyses. We allow evaluations of *auxiliary hypotheses* by including reliability of our measures (Table S2), which all exceeded reliability measures reported in the original instrument development (Wright et al., 2011). Given that we have met these six criteria, our results qualify

as “No signal - inconsistent” in LeBel et al.’s terminology. That is, our replication effect size 95% confidence intervals include 0 (Study 1: $r = -0.10$, 95% CI $[-0.34, 0.15]$; Study 2: $r = 0.04$, 95% CI $[-0.26, 0.34]$) but not the original effect size point estimates from Brady et al.’s adult studies ($r = -0.46$ and -0.61).

To compare results across all of the studies published thus far, we conducted an exploratory meta-analysis that combined our data with those of Brady et al. (2018) and Mongillo et al. (2019). This meta-analysis found that, across these six studies, the overall effect size is in the direction demonstrated by Brady et al. However, the calculated effect size is substantially smaller than the original studies and the credibility interval includes 0. The Bayesian analysis found anecdotal evidence for no between-study heterogeneity and no correlation between distance traveled and DIAS scores. But because the evidence is only anecdotal, we cannot draw strong conclusions based on these results. Moreover, we must acknowledge the assumptions and limitations of Bayesian meta-analyses. First, we used a single set of priors for the Bayesian model based on the defaults suggested for Fisher’s z scale by Heck et al. (2019). Other priors might result in different outcomes. Second, the sample size for this meta-analysis was relatively small with only six studies. More studies are needed to have better estimates of the effect sizes. Moreover, selection of studies is critical. For instance, we included Brady et al.’s Lab Study 2—which focused on puppies—to maximize the number of studies included in the analysis. But an argument could be made to exclude this study since it focused on puppies (for our purposes, removing this study does not change any of our findings). Finally, a key weakness of meta-analyses is that they are distorted by publication bias. Given the relatively recent publication date of Brady et al. (2018), it is unlikely that there are many additional studies at the moment that have been conducted but not published. Nevertheless, failures to replicate are not published at the same rate as confirming replications (Francis, 2012; Makel et al., 2012; Martin & Clarke, 2017), and the resulting publication bias may generate a larger overall effect size than is accurate (Simonsohn et al., 2014; Friesse & Frankenbach, 2020). In summary, the meta-analysis does not resolve the presence of an effect, so more studies are needed.

Why might we see such variability across studies? The first possibility is that we did not accurately replicate Brady et al. (2018). However, we made every effort to replicate the methods described in their article. Though our first study used a smaller space and maximum distance, we rectified this in the second study. Otherwise, our methods were very similar to the original study.

A second possibility is that we sampled a different subset of the dog population than the original study. One clear way that this could happen could be based on age. Brady et al.’s two

adult studies (that showed the effect) involved dogs that were 2 – 10 years old (mean \pm SD = 5.3 ± 2.3), whereas dogs in our studies ranged from 0.5 – 16 (mean \pm SD = 4.7 ± 3.2) years old. However, trimming our data down to 2 – 10 year old dogs ($N = 42$, mean \pm SD = 5.0 ± 2.4) did not change the results for the correlation between maximum distance traveled and DIAS overall score for the combined study data ($r = -0.05$, $BF_{10} = 0.27$), indicating that dog age differences across the samples do not account for effect differences.

Relatedly, there could be breed differences between study samples. From the tables in Brady et al. (2018), it appears as though the vast majority of dogs in their samples are purebred (5 of 37 or 13.5% are listed as ‘crossbreed’). In our samples, 37 out of 108 (34.2%) were listed as mixed breed (data set includes owners’ descriptions of breed). This is likely a low estimate, as we did not explicitly ask if the dogs were purebred. Breed composition matters because breeds may differ in their impulsivity (Fadel et al., 2016; Gerencsér et al., 2018), though this is not always the case (Lit et al., 2010). Moreover, owners may have expectations about breed differences in impulsivity. Therefore, differences in breed composition in samples may result in different relationships between owner reports and dog impulsivity.

A third possibility is that we sampled a different subset of the owner population than Brady et al. (2018). While demographic information about dogs is commonly provided in the dog behavior literature, information on the owners is often not. This is important because as researchers, we are selectively sampling from a subset of dog owners—those who have the time, funds, and interest to bring their dogs in for testing. These samples are not representative and may have critical influences on generalizability, so we need to encourage more thorough reporting of owner information to capture potential differences between studies. We provide summary information on the owners’ gender identity, marital status, and household income category in Table S1 and individual values in our data set. In hindsight, we would like to have collected information about age, employment status, and time spent at home with their dogs. Factors such as age, marital status, income, employment status, and time spent with the dog are important variables to consider when examining owner knowledge about their dogs and impulsivity specifically. If our lab sampled a different subset of owners than Brady et al., our sample may have different knowledge about their dogs’ impulsivity, which could drive the study differences. Alternatively, a different subsample of owners may select or train their dogs differently, which can result in more variability in impulsivity levels, again resulting in study differences. Given the critical importance of owners in raising and training the dogs in our samples, canine science researchers should improve and standardize the collection and availability of information about owners.

The dog and owner sample differences imply that our lab may simply draw a different subset of dogs and owners from a common population compared to the other studies. However, a fourth possibility is that there are real differences between the populations to draw from; that is, cultural differences may exist between our populations and those of Brady et al. (2018). Cultures differ in their attitudes toward and beliefs about dogs (Bradshaw & Goodwin, 1999; Serpell, 2004) and their experiences and interactions with dogs (Wan et al., 2009; Amici et al., 2019). Our study was completed in the United States (US), whereas Brady et al. was conducted in the United Kingdom (UK), two countries that differ in the frequency of dog ownership. In 2021, approximately 53% of US households had dogs (American Pet Products Association, 2021), compared to 33% of UK households (Pet Food Manufacturers' Association, 2021). The two countries also differ in many aspects of dog welfare, health care, attitudes, policies, and laws (Houpt et al., 2007). Though there are few direct comparisons of US and UK dog owners, there is some evidence that veterinarians' beliefs about breed characteristics are shared for some breeds but differ for other breeds between the two countries (Bradshaw & Goodwin, 1999). Differences in attitudes and beliefs about dogs could influence how owners train, relate to, and interact with their dogs. These differences in turn could influence both how impulsive their dogs are, as well as how well owners know their dogs' behavioral traits. For instance, US owners are more likely than UK owners to spay or neuter their dogs (Diesel et al., 2010; Trevejo et al., 2011). Because neutered male dogs may be more impulsive than non-neutered males (Fadel et al., 2016), different frequencies of neutered males may influence distributions of impulsivity scores. In our sample, 94.7% of the males were neutered. Though Brady et al. do not report their neuter rate, it was likely much lower than ours. However, UK and Italian owners share similar attitudes toward dogs (Lakestani et al., 2011), but Mongillo et al. (2019) did not replicate the effects of Brady et al. either. Thus, potential cultural differences between study populations should be further investigated.

Implications

Along with Mongillo et al. (2019), our results call into question the robustness of the relationship between performance on a spatial impulsivity task and an owner's report of dog impulsivity more generally. Four out of six studies on this topic do not show a relationship (Figure 4). A meta-analysis does not yield definitive evidence in either direction, though it provides anecdotal evidence favoring no relationship. What does this mean for canine impulsivity?

Brady et al. (2018) investigated the relationship between spatial impulsivity and DIAS because they were interested in developing a measure of impulsivity that could be assessed quickly; that is, it did not require the repeated interactions

needed to offer an accurate assessment of impulsivity via the DIAS. This could be useful especially in settings such as animal shelters, where quickly assessing dog impulsivity can trigger additional training or help find a match with potential adopters. Our study comparison suggests that the spatial impulsivity task may not provide a broadly reliable measure of impulsivity for these purposes. We discussed the possibility of cultural/country differences potentially explaining the divergent findings across studies. This implies that the spatial impulsivity task may be a useful general measure of impulsivity in some situations (e.g., UK households) but not others. To estimate how generalizable it is as a measure of impulsivity, the task should be replicated more both within and outside of the UK.

Not finding a relationship between spatial impulsivity and DIAS scores, however, is perhaps not surprising given other information about impulsivity. Impulsivity is a multi-faceted concept that refers to many different cognitive and motivational processes (Evenden, 1999; Reynolds et al., 2006; Stevens, 2017b). In humans, while there are some correlations across different measures of impulsivity, there are also many aspects of impulsivity that are unrelated (Whiteside & Lynam, 2001; Weber et al., 2002; DeYoung, 2011). Someone who is impulsive in the sense that they may gorge themselves on a decadent dessert may not be impulsive in the sense of engaging in unprotected sex. Similarly, the literature in dogs has failed to find relationships among different behavioral measures of impulsivity and inhibitory control, despite numerous studies searching for a consistent behavioral trait of impulsivity (Bray et al., 2014; Fagnani et al., 2016; Brucks, Marshall-Pescini, et al., 2017; van Horik et al., 2018). Thus, the fact that we found no noticeable relationship between spatial impulsivity and owner reports of impulsivity matches a broader set of results on this phenomenon across species.

The spatial impulsivity task likely taps into what would be considered *impulsive choice* rather than *impulsive action*, which refers to the inability to inhibit a response (Stevens, 2017b). Many of the classic inhibition tasks would fall under the impulsive action category (Bray et al., 2014; Brucks, Marshall-Pescini, et al., 2017; van Horik et al., 2018). Impulsive choice refers to situations in which individuals must choose between different reward amounts with different associated costs (e.g., intertemporal choice or risky choice, Stevens, 2017b). For spatial impulsivity, the cost is the time and energy required to obtain the larger rewards (Stevens et al., 2005). Having well-defined descriptions of concepts related to impulsivity may help clarify how concepts are related as well as help refine what types of methods best capture the concepts.

Given the distinction between impulsive choice and action, it is important to understand what aspects of impulsivity the DIAS captures. For instance, studies on impulsive action have not demonstrated relationships with DIAS (Mitcham,

2015; Brucks, Marshall-Pescini, et al., 2017; Cavalli et al., 2018). Research using measures of impulsive choice such as intertemporal choice (Wright et al., 2012; Riemer et al., 2014) and spatial impulsivity (Brady et al., 2018) have shown a relationship with DIAS; however, others have failed to show relationships in these types of impulsive choice tasks (Fagnani et al., 2016; Brucks, Soliani, et al., 2017; Mongillo et al., 2019). The items included in the behavioral regulation component of DIAS ask owners generally about whether their dogs are impulsive, are patient, are persistent, think before they act, demonstrate repetitive behaviors, have control over how they respond, and calm down quickly after getting excited. Interestingly, these items tend to be either rather general or vague terms (impulsive, patient) or refer to impulsive action (think before they act, have control over responses). Thus, one might not expect strong relationships with measures of impulsive choice, especially spatial impulsivity, which has minimal connection to the items in the DIAS. Without a stronger connection between DIAS and behavioral measures of impulsivity, it is difficult to use the DIAS or the impulsivity tasks in an applied setting, for instance, to make recommendations about using either measure as information for finding an appropriate home for a dog.

Challenges and opportunities

Both direct and conceptual replications can be difficult to implement. We applaud Brady et al. (2018) for facilitating replication by offering detailed descriptions of their methods and providing the individual data for their findings. Clear, thorough descriptions of methods are critical to providing the opportunity for others to replicate methods. We would also encourage the public posting of example videos of trials to further clarify methods. Additionally, individual-level data provided by Brady et al. facilitated our ability to compare the studies more directly in ways that cannot be compared with simple summary statistics. Publicly available trial-level data can be even more beneficial to others, as it offers more granular levels of analysis and more flexible statistical models such as generalized linear models.

We advocate for leveraging aspects of open science to further advance research on dog behavior and cognition (Stevens, 2017a; Beran, 2018). One of the first ways to improve the robustness of our findings is to recruit larger sample sizes. For correlational work in particular, small sample sizes can result in spurious relationships (Schönbrodt & Perugini, 2013; Loken & Gelman, 2017). While many experts are hesitant to offer more statistical rules of thumb, larger samples are better, and this is especially relevant in dogs given their variation in size and breed-specific selection differences. Carefully considering in advance what your sample size will be (Lakens, 2022) or using sequential hypothesis testing with Bayes factors as we do here (Schönbrodt et al., 2017) is critical to

collecting large enough samples to yield robust results. In addition, formally pre-registering or at least informally pre-specifying analysis plans is critical for reducing researcher degrees of freedom in analyzing data (Wagenmakers et al., 2012). For instance, including in the analysis the dog that Brady et al. (2018) excluded due to a diagnosis of hip dysplasia after the study dramatically changes the outcome of their analysis, dropping the correlation coefficient from $r = -0.46$ to $r = -0.37$ and increasing the p-value from 0.028 to 0.077. At small sample sizes, Pearson correlations are vulnerable to outliers and robust analyses are required (Wilcox, 2004). Flexibility in outlier exclusions combined with small sample sizes can reduce robustness of results, substantially influencing interpretations of our studies.

Producing a robust, reliable canine science requires implementing transparent and open research practices (Stevens, 2017a; Beran, 2018). One initiative that provides this opportunity in canine science is the ManyDogs Project (<http://manydogs.org/>). This project is a consortium of dog researchers around the globe interested conducting the same studies in all of their sites (ManyDogs et al., 2021). This not only provides much larger sample sizes than are reasonable at individual sites but also allows for direct comparison of similarities/differences across breeds and countries. Moreover, the project is devoted to open science practices including pre-registration of hypothesis, methods, and analysis plans, as well as public posting of data. Even if a researcher does not directly participate in it, the ManyDogs Project provides a model for transparent and open practices for canine science.

Finally, we raise an open secret in the dog cognition world. These studies were our lab's first research projects that involved bringing dogs into the lab for behavioral testing. One unexpected finding from this endeavor was how many dogs did not advance to the actual experiment. Out of 220 dogs that we recruited in total, 96 advanced to testing. A 56.4% failure-to-advance rate was unexpectedly high. While some of the dogs failed to advance because they were not treat motivated or failed to reliably prefer the larger reward in training, the vast majority of these dogs seemed to not engage in the task because they were too focused on their owner being out of the room. In Study 1, we intentionally kept owners in a different room to avoid cuing or interactions with the dog. However, many dogs seemed to exhibit separation anxiety with their owners out of the room. After noticing the high drop-out rate, we began asking owners if their dogs experienced separation anxiety. But informally, this did not seem to predict whether their dog would advance or not. We attempted to minimize this issue in Study 2 by starting with dogs who were already at the dog daycare and therefore should not experience separation anxiety. However, we only recruited a moderate number of dogs this way and recruited the remaining subjects through our database again.

We raise the issue of separation anxiety here to (1) be explicit about this happening and (2) highlight that the data are coming from an even smaller slice of the dog population than it may seem. Not only are we primarily recruiting a subset of people who are interested in and willing to come in to do the testing, but we are collecting data from a subset of those whose dogs are comfortable being in a different room from their owners. Thus, our findings are not necessarily representative of dogs generally, and this could potentially be another population difference between our sample and that of Brady et al. (2018). Furthermore, the recent COVID-19 pandemic has resulted in people staying home more, which could potentially influence separation anxiety in dogs (Heirs & Graham, 2021; Harvey et al., 2022). Further work is needed to investigate this, but it is important in considering recruitment and interpretation of dog cognition research. Going forward, we recommend allowing owners in the testing room with their dogs to reduce the drop-out rate. This requires additional thought and planning about ways to avoid owners intentionally and unintentionally cuing their dogs. However, the potential benefits of higher success in data collection will likely outweigh the costs of additional time and thought required to minimize cuing effects.

Conclusion

Impulsivity is a critical feature of canine science. Finding easy measures of impulsivity could benefit dog training, obedience, welfare, breeding, and re-homing. However, impulsivity is a multi-faceted construct, and being impulsive in one context may not predict impulsivity in another context. We found that impulsivity in a spatial context did not relate to the owner's overall assessment of a dog's impulsivity. While this did not replicate the findings of Brady et al. (2018), it did match Mongillo et al.'s (2019) finding of no relationship. There are many potential reasons for our inability to replicate Brady et al., ranging from potential methodological and population differences to larger cultural differences between sites. This outcome highlights the importance of implementing robust open science practices across canine science to help translate our results into practice in ways that benefit both owners and dogs.

Acknowledgments

This research was funded in part by the National Science Foundation (NSF-1658837) and by private funds donated to the University of Nebraska Foundation. We would like to thank Dian Quist from Kenl Inn for allowing us to use her facilities and recruit from the dogs in her dog daycare. We also thank the Kenl Inn staff for helping bring the dogs to the testing area. We are grateful to the following assistants for helping to test the dogs: Jessica Barela, Meredith Batten, Toria Biancalana, Katie Carey, Hunter DeBoer, Rose Felice,

Haley Hays, Billy Lim, Brianna Moser, McKenna Rezny, Joelle Sanger, Taylor Schendt, and Destiny Vail.

Author Contributions

Stevens: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Visualization, Writing – Original Draft Preparation. **Mathias:** Investigation, Project Administration, Supervision, Writing – Review & Editing. **Herridge:** Investigation, Project Administration, Supervision, Writing – Review & Editing. **Hughes-Duvall:** Investigation, Project Administration, Supervision, Writing – Review & Editing. **Wolff:** Investigation, Supervision, Writing – review & editing. **Yohe:** Investigation, Writing – Review & Editing.

Conflict of interest: The authors declared that no conflicts of interest exist.

Data Availability: The data and analysis code are available at: <https://doi.org/10.17605/osf.io/eb5m3>.

References

- American Pet Products Association. (2021). 2021-2022 APAA National Pet Owners Survey. https://americanpetproducts.org/pubs_survey.asp.
- Amici, F., Waterman, J., Kellermann, C. M., Karimullah, K., & Bräuer, J. (2019). The ability to recognize dog emotions depends on the cultural milieu in which we grow up. *Scientific Reports*, 9(1), 1–9. <https://doi.org/10.1038/s41598-019-52938-4>
- Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, 41(2), 521–543. <https://doi.org/10.1177/0149206314560412>
- Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://github.com/crsh/papaja>
- Bennett, P. C., & Rohlf, V. I. (2007). Owner-companion dog interactions: Relationships between demographic variables, potentially problematic behaviours, training engagement and shared activities. *Applied Animal Behaviour Science*, 102(1), 65–84. <https://doi.org/10.1016/j.applanim.2006.03.009>
- Beran, M. (2018). Replication and pre-registration in comparative psychology. *International Journal of Comparative Psychology*, 31(0).
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (2nd ed.). John Wiley & Sons.
- Bradshaw, J. W. S., & Goodwin, D. (1999). Determination of behavioural traits of pure-bred dogs using factor analysis and cluster analysis; a comparison of studies in the USA and UK. *Research in Veterinary Science*, 66(1), 73–76. <https://doi.org/10.1053/rvsc.1998.0246>
- Brady, K., Hewison, L., Wright, H., Zulch, H., Cracknell, N., & Mills, D. (2018). A spatial discounting test to assess impulsivity in dogs. *Applied Animal Behaviour Science*, 202, 77–84. <https://doi.org/10.1016/j.applanim.2018.01.003>
- Bray, E. E., MacLean, E. L., & Hare, B. A. (2014). Context specificity of inhibitory control in dogs. *Animal Cognition*, 17(1), 15–31. <https://doi.org/10.1007/s10071-013-0633-z>
- Brucks, D., Marshall-Pescini, S., Wallis, L. J., Huber, L., & Range, F. (2017). Measures of dogs' inhibitory control abilities do not correlate across tasks. *Frontiers in Psychology*, 8, 849. <https://doi.org/10.3389/fpsyg.2017.00849>
- Brucks, D., Soliani, M., Range, F., & Marshall-Pescini, S. (2017). Reward type and behavioural patterns predict dogs' success in a delay of gratification paradigm. *Scientific Reports*, 7, 42459. <https://doi.org/10.1038/srep42459>
- Cavalli, C. M., Carballo, F., Dzik, M. V., Underwood, S., & Bentosela, M. (2018). Are animal-assisted activity dogs different from pet dogs? A comparison of their sociocognitive abilities. *Journal of Veterinary Behavior*, 23, 76–81. <https://doi.org/10.1016/j.jveb.2017.12.001>
- Clarke, E., & Sherrill-Mix, S. (2017). *ggbeeswarm: Categorical scatter (violin point) plots*. <https://CRAN.R-project.org/package=ggbeeswarm>
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, 7(1), 25–47.

- DeYoung, C. G. (2011). Impulsivity as a personality trait. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications*, 2nd ed (pp. 485–502). Guilford Press.
- Diesel, G., Brodbelt, D., & Laurence, C. (2010). Survey of veterinary practice policies and opinions on neutering dogs. *Veterinary Record*, 166(15), 455–458. <https://doi.org/10.1136/vr.b4798>
- Dwyer, F., Bennett, P. C., & Coleman, G. J. (2006). Development of the Monash Dog Owner Relationship Scale (MDORS). *Anthrozoös*, 19(3), 243–256. <https://doi.org/10.2752/089279306785415592>
- Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology*, 146(4), 348–361. <https://doi.org/10.1007/PL00005481>
- Fadel, F. R., Driscoll, P., Pilot, M., Wright, H., Zulch, H., & Mills, D. (2016). Differences in trait impulsivity indicate diversification of dog breeds into working and show lines. *Scientific Reports*, 6(1), 22162. <https://doi.org/10.1038/srep22162>
- Fagnani, J., Barrera, G., Carballo, F., & Bentosela, M. (2016). Is previous experience important for inhibitory control? A comparison between shelter and pet dogs in A-not-B and cylinder tasks. *Animal Cognition*, 19(6), 1165–1172. <https://doi.org/10.1007/s10071-016-1024-z>
- Fawcett, T. W., McNamara, J. M., & Houston, A. I. (2012). When is it adaptive to be patient? A general framework for evaluating delayed rewards. *Behavioural Processes*, 89(2), 128–136. <https://doi.org/10.1016/j.beproc.2011.08.015>
- Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, 19(6), 975–991. <https://doi.org/10.3758/s13423-012-0322-y>
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25–42. <https://doi.org/10.1257/089533005775196732>
- Friese, M., & Frankenbach, J. (2020). P-hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, 25(4), 456–471. <https://doi.org/10.1037/met0000246>
- Gerencsér, L., Bunford, N., Moesta, A., & Miklósi, Á. (2018). Development and validation of the Canine Reward Responsiveness Scale—Examining individual differences in reward responsiveness of the domestic dog. *Scientific Reports*, 8(1), 4421. <https://doi.org/10.1038/s41598-018-22605-1>
- Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, 37(6), 504–528. [https://doi.org/10.1016/S0092-6566\(03\)00046-1](https://doi.org/10.1016/S0092-6566(03)00046-1)
- Gronau, Q. F., Heck, D. W., Berkhout, S. W., Haaf, J. M., & Wagenmakers, E.-J. (2021). A primer on Bayesian model-averaged meta-analysis. *Advances in Methods and Practices in Psychological Science*, 4(3), 25152459211031256. <https://doi.org/10.1177/25152459211031256>
- Harvey, N. D., Christley, R. M., Giragosian, K., Mead, R., Murray, J. K., Samet, L., Upjohn, M. M., & Casey, R. A. (2022). Impact of changes in time left alone on separation-related behaviour in UK pet dogs. *Animals*, 12(4), 482. <https://doi.org/10.3390/ani12040482>
- Heck, W. D., Gronau, F. Q., Wagenmakers, E.-J., & E.-J. (2019). *metaBMA: Bayesian model averaging for random and fixed effects meta-analysis*. <https://CRAN.R-project.org/package=metaBMA>
- Heirs, M., & Graham, H. (2021). Canine separation anxiety: How first referral practices can help. *Companion Animal*, 26(8), 170–174. <https://doi.org/10.12968/coan.2021.0030>
- Hejjas, K., Vas, J., Topal, J., Szantai, E., Ronai, Z., Szekely, A., Kubinyi, E., Horvath, Z., Sasvari-Szekely, M., & Miklósi, A. (2007). Association of polymorphisms in the dopamine D4 receptor gene and the activity-impulsivity endophenotype in dogs. *Animal Genetics*, 38(6), 629–633. <https://doi.org/10.1111/j.1365-2052.2007.01657.x>
- Hiby, E., Rooney, N., & Bradshaw, J. (2004). Dog training methods: Their use, effectiveness and interaction with behaviour and welfare. *Animal Welfare*, 13(1), 63–69.
- Hopper, L. M., Kurtycz, L. M., Ross, S. R., & Bonnie, K. E. (2015). Captive chimpanzee foraging in a social setting: A test of problem solving, flexibility, and spatial discounting. *PeerJ*, 3, e833. <https://doi.org/10.7717/peerj.833>
- Houpt, K. A., Goodwin, D., Uchida, Y., Baranyiová, E., Fatjó, J., & Kakuma, Y. (2007). Proceedings of a workshop to identify dog welfare issues in the US, Japan, Czech Republic, Spain and the UK. *Applied Animal Behaviour Science*, 106(4), 221–233. <https://doi.org/10.1016/j.applanim.2007.01.005>
- 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>
- Kay, M. (2022). *ggdist: Visualizations of distributions and uncertainty*. <https://doi.org/10.5281/zenodo.3879620>
- Kralik, J. D., & Sampson, W. W. L. (2012). A fruit in hand is worth many more in the bush: Steep spatial discounting by free-ranging rhesus macaques (*Macaca mulatta*). *Behavioural Processes*, 89(3), 197–202. <https://doi.org/10.1016/j.beproc.2011.09.010>
- Kubinyi, E., Vas, J., Hejjas, K., Ronai, Z., Brúder, I., Turcsán, B., Sasvari-Szekely, M., & Miklósi, Á. (2012). Polymorphism in the tyrosine hydroxylase (TH) gene is associated with activity-impulsivity in German Shepherd Dogs. *PLOS ONE*, 7(1), e30271. <https://doi.org/10.1371/journal.pone.0030271>
- Lakens, D. (2022). Sample size justification. *Collabra: Psychology*, 8(1), 33267. <https://doi.org/10.1525/collabra.33267>
- Lakestani, N., Donaldson, M. L., Verga, M., & Waran, N. (2011). Attitudes of children and adults to dogs in Italy, Spain, and the United Kingdom. *Journal of Veterinary Behavior*, 6(2), 121–129. <https://doi.org/10.1016/j.jvbeh.2010.11.002>
- LeBel, E. P., Vanpaemel, W., Cheung, I., & Campbell, L. (2019). A brief guide to evaluate replications. *Meta-Psychology*, 3, MP.2018.843. <https://doi.org/10.15626/MP.2018.843>
- Lit, L., Schweitzer, J. B., Iosif, A.-M., & Oberbauer, A. M. (2010). Owner reports of attention, activity, and impulsivity in dogs: A replication study. *Behavioral and Brain Functions*, 6(1), 1. <https://doi.org/10.1186/1744-9081-6-1>
- Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, 355(6325), 584–585. <https://doi.org/10.1126/science.aal3618>
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542. <https://doi.org/10.1177/1745691612460688>
- ManyDogs, Espinosa, J., Bray, E., Buchsbaum, D., Byosiore, S.-E., Byrne, M., Freeman, M. S., Gnanadesikan, G., Guran, C.-N. A., Horschler, D., Huber, L., Johnston, A., MacLean, E., Pelgrim, M. H., Santos, L., Silver, Z. A., Stevens, J. R., Völter, C., & Zipperling, L. (2021). *ManyDogs 1: A multi-lab replication study of dogs' pointing comprehension*. <https://doi.org/10.31234/osf.io/f86jq>
- Martin, G. N., & Clarke, R. M. (2017). Are psychology journals anti-replication? A snapshot of editorial practices. *Frontiers in Psychology*, 8, 523. <https://doi.org/10.3389/fpsyg.2017.00523>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70(6), 487–498. <https://doi.org/10.1037/a0039400>
- McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, 23(3), 412–433. <https://doi.org/10.1037/met0000144>
- Mitcham, (2015). *Effect of Socialization on Impulsivity in the Domestic Dog* [Master's thesis]. University of Guelph.
- Mongillo, P., Scandurra, A., Eatherington, C. J., D'Aniello, B., & Marinelli, L. (2019). Development of a spatial discount task to measure impulsive choices in dogs. *Animals*, 9(7), 469. <https://doi.org/10.3390/ani9070469>
- Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of Bayes factors for common designs*. <https://CRAN.R-project.org/package=BayesFactor>
- Mühlhoff, N., Stevens, J. R., & Reader, S. M. (2011). Spatial discounting of food and social rewards in guppies (*Poecilia reticulata*). *Frontiers in Psychology*, 2, 68. <https://doi.org/10.3389/fpsyg.2011.00068>
- Müller, K. (2020). *here: A simpler way to find your files*. <https://CRAN.R-project.org/package=here>
- Papale, A. E., Stott, J. J., Powell, N. J., Regier, P. S., & Redish, A. D. (2012). Interactions between deliberation and delay-discounting in rats. *Cognitive, Affective, & Behavioral Neuroscience*, 12(3), 513–526. <https://doi.org/10.3758/s13415-012-0097-7>
- Pedersen, T. L. (2020). *patchwork: The composer of plots*. <https://CRAN.R-project.org/package=patchwork>
- Pet Food Manufacturers' Association. (2021). *Pet Population 2021*. <https://www.pfma.org.uk/pet-population-2021>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University. <https://CRAN.R-project.org/package=psych>
- Revelle, W., & Zinbarg, R. E. (2008). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145–154. <https://doi.org/10.1007/s11336-008-9102-z>
- Reynolds, B., Ortengren, A., Richards, J. B., & de Wit, H. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, 40(2), 305–315. <https://doi.org/10.1016/j.paid.2005.03.024>
- Riemer, S., Mills, D. S., & Wright, H. (2014). Impulsive for life? The nature of long-term impulsivity in domestic dogs. *Animal Cognition*, 17(3), 815–819. <https://doi.org/10.1007/s10071-013-0701-4>
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21(2), 301–308. <https://doi.org/10.3758/s13423-014-0595-4>
- Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609–612. <https://doi.org/10.1016/j.jrp.2013.05.009>
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), 322–339. <https://doi.org/10.1037/met0000061>
- Serpell, J. A. (2004). Factors influencing human attitudes to animals and their welfare. *Animal Welfare*, 13(S1), 145–151.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666–681. <https://doi.org/10.1177/1745691614553988>
- Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: Are yours realistic? *Perspectives on Psychological Science*, 9(3), 305–318. <https://doi.org/10.1177/1745691614528518>
- Stevens, J. R. (2017a). Replicability and reproducibility in comparative psychology. *Frontiers in Psychology*, 8, 862. <https://doi.org/10.3389/fpsyg.2017.00862>
- Stevens, J. R. (2017b). The many faces of impulsivity. In J. R. Stevens (Ed.), *Impulsivity: How Time and Risk Influence Decision Making* (pp. 1–6). Springer International