Supporting Information

Bray et al. 10.1073/pnas.1704303114

SI Materials and Methods

Breed Classifications. Puppies from cross-litters ($n = 31$) were categorized as the breed that contributed over 50% of their genes. If they were 50% Labrador Retriever and 50% Golden Retriever, they were assigned their mother's breed (Table S1). In total, we tested 51 German Shepherds, 60 Labrador Retrievers, and 22 Golden Retrievers. In the final analysis of program outcome ($n = 98$), the sample consisted of 39 German Shepherds, 44 Labrador Retrievers, and 15 Golden Retrievers (Table S3).

***Material Behavior*.** Mothers were housed singly in indoor pens with access to an outdoor area through a guillotine door, and puppies were contained in towel-lined kiddie pools ("nursing boxes") over the first 3 wk. We coded distinct behaviors by mothers: time spent in the nursing box with puppies, contact, licking/grooming, lateral nursing (mother lying on side), vertical nursing (mother sitting/standing), ventral nursing (mother lying on stomach), and orienting away from puppies. These behaviors all loaded strongly onto one PC, *Maternal behavior*. This component explained a significant portion of the variance, remained stable over time, and had predictive validity because it was correlated with independent experimental and hormonal measures of mothering (36). Mothers that scored high on *Maternal behavior* were vigilant; often in proximity to their litter; and regularly contacted, licked, groomed, and nursed their puppies.

***Maternal Behavior* as a Predictor Variable.** Because *Maternal behavior* scores were significantly positively correlated across weeks 1, 2, and 3 (36), we use week 2 *Maternal behavior* scores in all analyses as our predictor variable. We were unable to observe two litters on week 2, so these puppies were given the average of their mothers' week 1 and week 3 *Maternal behavior* scores. Because we could not identify puppies individually on our videotapes, all littermates received the same score.

***Maternal Behavior* Association with Outcome.** We conducted a GEE-GLM with outcome as the dependent variable, *Maternal behavior* as the predictor variable, and litter ID as a random effect. Breed, maternal parity, sex of puppy, and age at return were included as covariates. Results (Table S5) revealed a significant main effect of *Maternal behavior* (Wald = 12.98, $P < 0.001$): Puppies raised by mothers exhibiting more maternal behavior were more likely to be released from the program (OR = 3.39). The odds of program release were 3.39-fold higher with each SD increase in *Maternal behavior*. None of the covariates were significant predictors.

Testing After Return to The Seeing Eye and After Surgery. After returning to headquarters for training, males were housed individually, whereas females were often housed with a same-sex kennelmate. All dogs received food twice a day (7:00 hours and 16:00 hours), and water was always available. The lights were switched on around 6:30 hours in the morning and turned off at 18:00 hours at night.

Young adult testing took place from May through October 2015. Most dogs completed postarrival (PA) testing 2 d ($n = 107$) or 3 d ($n = 22$) after they were returned to headquarters from their puppy-raising families. One dog was tested 1 d PA, two were tested 5 d PA, and one was tested 6 d PA. The majority then completed a second round of postsurgery (PS) testing 2 d ($n = 100$) or 3 d ($n = 24$) after undergoing anesthesia for alteration surgery and/or hip X-rays. One dog each was tested 8, 16, 21, 22, and 23 d PS, whereas four dogs did not undergo anesthesia at all before PS testing. PS testing was identical to PA testing.

All testing occurred in an 11 × 7-ft examination room located within an unoccupied kennel wing. Testing occurred between 7:30 hours and 17:30 hours. On a testing day, each dog was tested twice and tasks were always presented in the same order. Within a given dog's testing session, tasks occurred one after the other with only brief breaks for setup. Food rewards consisted of Zuke's Mini Naturals treats. Testing sessions were videotaped using Sony video cameras (HDR-PJ230, HDR-CX405) mounted on tripods.

The following variables were coded from videotape either by one of the authors (E.E.B.) or by a research assistant who had participated in data collection: all variables from isolation, perseveration during multistep problem solving, all variables from the novel object task, and initial response during the umbrella-opening task. To assess the reliability of the video-coded variables, an additional coder coded 20% of randomly selected trials. The interrater reliability was assessed by calculating Spearman's rho for continuous variables and Cohen's kappa for categorical variables (Table S2). All other performance measures were coded live.

Data Reduction Applied to Young Adult Cognitive and Temperament Tasks. To determine which variables to include in the young adult analysis, we compared the dogs' rankings on the 29 scores (derived from 11 tasks; Table S2) at PA testing with their rankings on these same scores at PS testing by computing Kendall rank correlation coefficients. Results revealed significant correlations ($P < 0.05$) for 21 of 29 scores and a marginally significant correlation ($P = 0.09$) for one of 29 scores. Seven of 29 scores were not significantly correlated ($P > 0.10$). Thus, the dogs' rank order of performance was significantly correlated in 21 of 29 (72%) task variables. Moreover, the lack of correlation in three scores (cylinder task and two detour problem-solving performance measures) was most likely due to a ceiling effect on the task at the time of the second testing. Finally, the conditions surrounding PA testing were much more consistent across dogs than the conditions surrounding PS testing. PA testing happened 2 to 3 d PA in 97% of subjects. PS testing happened 2 to 3 d PS in 92% of subjects, but time of surgery varied by dog and ranged from 1 to 30 d PA. In addition, some dogs never had surgery [e.g., dogs earmarked for the breeding program at the time of testing ($n = 15$), dogs that had previously been spayed or neutered ($n = 4$), dogs that had medical issues ($n = 2$)], and three of the four dogs that were previously altered did not undergo anesthesia before their second testing. We therefore elected to use only results from PA (the initial) testing in subsequent analyses. Only seven dogs (five males and two females) were altered before young adult testing, so data for altered and intact dogs were not considered separately.

Using PA data only, we looked at each of the 11 tasks to determine how and/or if each of the task variables could be summarized in one or two PCs. Results revealed that the 29 variables from all tasks could be summarized by 15 measures: 13 PCs using a varimax rotation and two z-scored variables that were not appropriate for PCA due to unacceptable Kaiser–Meyer–Olkin values below 0.50 (Table S7).

Given the modest size of our dataset, it was necessary to reduce the total number of young adult behaviors that could be considered in a multivariate model (33). To screen the young adult

behaviors most associated with outcome in the program, we first evaluated each task score component in a separate GEE-GLM that clustered young adult dogs by litter as the unit of observation (Table S8). These models allowed us to evaluate the association between individual task score and program outcome after adjustment for important confounders (breed, maternal parity, sex of puppy, and age at return). We did not adjust for multiple testing because we expected that our task performance measures were correlated with one another, especially measures that were derived from the same task. Furthermore, because our goal was prediction, we needed to look at each measure's individual association with outcome to select the best smaller subset to be considered jointly (53).

Young Adult Testing Association with Outcome. The following scores were associated with release from the program: Dogs with slow solve times and high levels of perseveration on the multistep problem-solving task were more likely to be released (Table S9, task 5: OR = 1.75, Wald = 5.63, $P = 0.02$, odds of program release 1.75-fold higher; Movie S1), whereas dogs displaying a long latency to vocalize during the novel object task were less likely to be released (task 10b: OR = 0.43, Wald = 6.51, $P = 0.01$, odds of program release 57% lower; Movie S2). There was also a significant interaction between breed and reactivity in the umbrella-opening task (task 11a: OR = 0.23, Wald = 10.30, $P = 0.001$): Golden Retrievers that visibly reacted to the surprising event were less likely to be released from the program (Wald = 6.22, OR = 0.40, $P = 0.01$, odds of program release 60% lower), whereas Labrador Retrievers that reacted strongly were more likely to be released (Wald = 4.15, OR = 1.70, $P = 0.04$, odds of program release 1.70-fold higher). Age at return and sex were also significant independent predictors of outcome: Dogs that returned to headquarters at older ages had a lower probability of being released (Wald = 4.41, $P = 0.04$, OR = 0.59, odds of program release 41% lower), and the risk of program release was 69% lower for males than for females (Wald = 4.03, $P = 0.045$, OR = 0.31).

Discrimination of Models. To assess the discrimination of each model, we computed the areas under a receiver operating characteristic (ROC) curve using the R package "pROC" (35). Ninety-five percent confidence intervals were computed using the bootstrapping method, also in pROC (36). We then used paired-design "roc.tests" to compare the area under the ROC curve (AUC) of the different models with one another.

Associations Between *Maternal Behavior* and Measures of Young Adult Test Performance in Which There Was an Interaction Between *Maternal Behavior* and Breed. There was an interaction between breed and *Maternal behavior* on superior performance during memory problem solving (Table S4, task 4a: Wald = 4.68, $P = 0.03$), as well as between breed and *Maternal behavior* on accuracy during this task (task 4b; Wald = 8.81, $P = 0.003$): Golden Retrievers that experienced high levels of maternal care were persistent and quick to solve the memory problem-solving task (Wald = 6.25, $P = 0.01$), but they were also less accurate (Wald = 10.10, $P < 0.001$). We also found a significant interaction between breed and *Maternal behavior* on wariness during the novel object task (task 10a: Wald = 8.25, $P = 0.004$): Labrador Retrievers that experienced higher levels of maternal care oriented toward the novel objects at high levels but were slow to approach them (Wald = 9.16, $P < 0.01$). Finally, we found an interaction between breed and *Maternal behavior* on recovery during umbrella opening (task 11b; Wald = 7.57, $P < 0.01$): Labrador Retrievers that experienced higher levels of maternal behavior were slower to approach the umbrella postopening and spent less time in contact with it (Wald = 8.51, $P < 0.01$).

Table S1. Demographics of mothers and puppies in the study

Litter	Litter size	Puppies included in analyses	Mother's breed	Father's breed	Puppies' breed	Labrador Retriever, %	Coded puppy breed
Della	6	5	Labrador Retriever	Golden Retriever	Lab-Golden cross	50	Labrador Retriever
Lizzie	9	5	Golden Retriever	Golden Retriever	Golden Retriever	0	Golden Retriever
Dagmar	8	6	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Dori	5	2	Golden Retriever	Labrador Retriever	Lab-Golden cross	50	Golden Retriever
Lolly	2	0	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Dotty	2	1	Golden Retriever	Labrador Retriever	Lab-Golden cross	50	Golden Retriever
Onyx	8	5	Labrador Retriever	Labrador Retriever	Labrador Retriever	100	Labrador Retriever
Maude	9	5	Labrador Retriever	Labrador Retriever	Labrador Retriever	100	Labrador Retriever
Ayesha	10	7	Labrador Retriever*	Labrador Retriever	Lab-Golden cross × 3	87.5	Labrador Retriever
Foxy	7	5	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Toffee	6	5	Labrador Retriever	Labrador Retriever	Labrador Retriever	100	Labrador Retriever
Carey	8	5	Labrador Retriever	Labrador Retriever	Labrador Retriever	100	Labrador Retriever
Aura	7	6	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Naomi	8	6	Labrador Retriever*	Labrador Retriever	Lab-Golden cross × 3	87.5	Labrador Retriever
Omega	8	7	Golden Retriever	Golden Retriever	Golden Retriever	0	Golden Retriever
Lea	6	6	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Leah	5	3	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Paris	4	2	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Elise	9	8	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd
Xyris	7	6	Labrador Retriever	Labrador Retriever	Labrador Retriever	100	Labrador Retriever
Lisa	4	3	German Shepherd	German Shepherd	German Shepherd	0	German Shepherd

*These dogs are Labrador-Golden crosses × 2, meaning their mothers were 50%-50% Labrador-Golden crosses and their sires were 100% Labrador Retrievers, making them 75% Labrador Retriever. Thus, these dogs were classified as Labrador Retrievers.

Table S2. Summary of young adult tasks that were used in analysis

Order	Task	Task description	Variable	Type	Measure	Description	Rho	Kappa
1	Isolation	The handler releases the dog into the empty lighted testing room, which the dog is then free to explore for 2 min.	Time near exit	Duration	Time, %	Dog is near the exit, in the half of the room closest to the door	0.99	
			Activity score	Count	1–39	How many times dog switches between quadrants over the course of the session	0.98	
			Mobile	Duration	Time, %	Dog is not sitting, standing, or lying in the same spot for more than 3 s while in view	0.97	
			Vocalizing	Duration	Time, %	Dog is howling, barking, yelping, whining, groaning, or play-growling	0.94	
2	Distraction	The handler walks the dog to the end of a hallway, facing the experimenter, and releases the dog when the experimenter calls. During the first two trials, the hallway is empty. During the last two trials, six toys and three treat rewards are placed in exact, alternating locations. All trials are capped at 2 min.	Competency	Duration	No. of seconds	Amount of time to come when called to the experimenter down an empty hallway (44 × 4 ft), averaged over two trials		
			Toy distraction	Difference score		Amount of time to come when called down hallway with six toy and three food distractors minus amount of time to come when called down empty hallway		
			Toy contact	Average	0–6	Average number of toy distractors that dog contacts with any part of its body over two trials		
			Food ate	Average	0–3	Average number of treats that dog eats off of the floor of the hallway over two trials		
3	Sustained attention	The handler positions the dog to face the experimenter in a standing position. At the start of the trial, the experimenter says "[Dog's name], sit!" and holds up her right arm with a closed fist in a sit gesture, at which point the handler drops the leash. The trial begins when the dogs sits, and ends when both the dog's chest and face are oriented away from the experimenter. Each of two test trials is capped at 2 min.	Body orient, trial 1	Duration	No. of seconds	From the time the dog sits to the time that both the dog's chest and face are oriented away from the experimenter		
			Body orient, trial 2	Duration	No. of seconds	Same as above, for trial 2		
			Face orient, trial 1	Duration	No. of seconds	Time that dog's face is oriented toward the experimenter		
			Face orient, trial 2	Duration	No. of seconds	Same as above, for trial 2		
4	Memory problem solving	Over two stages of familiarization trials, dogs eat food treats directly out of food wells, as well as by removing plastic bones to uncover the wells as part of the Nina Ottosson Dog Magic game. In the test trial, the dog watches as the experimenter baits four of the nine wells with food and then places plastic bones over all of the wells. The dog is then released and allowed 2 min to solve the problem and retrieve the rewards.	Solving time	Duration	No. of seconds	Amount of time to uncover and eat all four treats successfully		
			No. correct	Count	0–4	Number of correct wells uncovered in 2 min		
			Accuracy score	Difference score	Correct wells, %	Number of correct wells uncovered in 2 min minus number of incorrect wells uncovered in 2 min		
			Persistence	Duration	Time, %	Amount of time engaging with the apparatus divided by the solving time		

Table S2. Cont.

Order	Task	Task description	Variable	Type	Measure	Description	Rho	Kappa
5	Multistep problem solving	Over three stages of familiarization trials, dogs eat treats directly out of uncovered wells, as well as spinning the apparatus and removing plastic bones to uncover the treats as part of the Nina Ottosson Dog Tornado game. In the test trial, the dog watches as the experimenter baits a well, twists the apparatus to cover the baited well, and then places a plastic bone in the empty well next to the baited one, thereby rendering the apparatus unable to spin until the bone is dislodged. The dog is released and allowed 2 min to solve the problem and retrieve the reward.	Solving time	Duration	No. of seconds	Amount of time to uncover and eat the treat successfully		
			Perseveration	Duration	Time, %	Amount of time that the dog sniffs, noses, paws, scratches, mouths, and/or licks at the area of the apparatus covering the well with the hidden treat while the bone is still in the adjacent well (and thus blocking the rotation of the apparatus), divided by the total amount of time interacting with the apparatus	0.99	
6	Cylinder	The dog completes familiarization trials with an opaque cylinder, in which it must retrieve a food reward from the open sides of the apparatus without touching the front on four of the last five trials. In 10 test trials, the dog must solve the identical problem except that the apparatus is a transparent cylinder.	Test trial score	Count	No. correct	Correct if dog's snout enters the open end of the cylinder without the dog first touching the exterior of the cylinder with any part of its snout or paws; incorrect if dog touches the front or back of the cylinder with its snout or paws before finding the treat		
7	Detour problem solving	The experimenter stands directly in front of the dog with a treat and calls it over three warm-up trials. Then, the handler takes the dog out of the room and the experimenter sets up a serpentine maze of barriers that are 4 ft tall. When the dog reenters the room, the experimenter stands out of view at the end of the maze and calls it, at which point the handler releases the dog to solve the problem. Three test trials are capped at 2 min, and if the dog has not solved the problem on its own by the end of each trial, it is shown the solution.	Test trial 1 time	Duration	No. of seconds	Amount of time from start of trial 1 to solving trial 1		
			Test trial 2 time	Duration	No. of seconds	Same as above, for trial 2		
			Test trial 3 time	Duration	No. of seconds	Same as above, for trial 3		
			Test trial score	Rating	1	Solved one of three trials within the time limit		
					2	Solved two of three trials within the time limit		
					3	Solved three of three trials within the time limit		
8	Greeting	The dog is held on leash in the center of the testing room. The experimenter knocks on the door, and then enters the room cloaked in a hooded felt cape and standing hunched-back, ~5 ft from the dog. She waits silently for 15 s, and then calls and encourages the dog. If the dog approaches, the experimenter pets the dog and talks in a friendly manner. At the end of 45 s, the experimenter removes her cape and plays with the dog.	Latency to approach	Latency	No. of seconds	Amount of time for dog to approach the experimenter after her entry into the testing room, with dogs that never approach receiving the maximum score of 45		
			Interact	Duration	No. of seconds	Amount of time that dog is in contact with the experimenter throughout the trial		

Table S2. Cont.

Order	Task	Task description	Variable	Type	Measure	Description	Rho	Kappa
9	Ball play	The experimenter throws a round rubber KONG extreme ball (medium/large) for 30 s as a warm-up, and then throws the ball and encourages the dog to retrieve it. She rethrows the ball as many times as the dog brings it back over 1 min and then repeats the process for a second trial.	Retrieval score	Rating	1	Dog shows no interest in the ball		
					2	Dog runs after the ball, touches it, but does not pick it up in its mouth		
					3	Dog picks up the ball but does not bring it back		
					4	Dog retrieves the ball and brings it back one to two times		
					5	Dog retrieves the ball and brings it back three or more times		
10	Novel object	The handler releases the dog into the empty testing room with two motion-activated battery-operated toy cats (FurReal Friends Daisy Play-With-Me-Kitty) for 2 min.	Latency to approach	Latency	No. of seconds	Amount of time to approach one of the cats initially within one foot	0.96	
			Orient	Duration	No. of seconds	Amount of time that dog spends with face oriented toward a cat	0.97	
			Latency to vocalize	Latency	No. of seconds	Amount of time until dog makes first sound (howl, bark, yelp, whine, groan, or play-growl)	0.84	
11	Umbrella opening	The handler holds the dog on leash 64-in from the experimenter. When the dog is facing forward, the experimenter pushes a button to release an auto-open black umbrella, then immediately lowers it to the ground. The dog is then allowed to explore for 45 s. If the dog is not near the umbrella after 15 s, the experimenter verbally coaxes it, and if the dog does not approach after 30 s, the handler will pick up the dog's tab leash and try to gently guide the dog to the umbrella.	Reactivity initial response	Rating	1	No detectable reaction other than turning head or perking ears		0.83
					2	Flinch or startle without lowering of the body (some movement, including a small step back, is fine)		
					3	Crouch or ducking (downward movement of body and/or head) without major displacement and maintaining general body orientation		
					4	Rapid avoidance response away from stimulus		
			Recovery approach	Rating	1	Dog initially approaches the umbrella within 15 s		
					2	Dog initially approaches the umbrella within 16–30 s, after receiving verbal encouragement		
					3	Dog initially approaches the umbrella after 30 s, after being led to it on leash		
					4	Dog never approaches the umbrella over the 45-s trial, despite verbal and physical coaxing		
			Recovery contact	Duration	No. of seconds	Dog closely sniffing and/or in contact with the umbrella		

The tasks presented during young adult testing were similar to the tasks reported in the following studies: task 1 (22, 37, 38), task 2 (26, 39), task 3 (not previously studied), task 4 (40), task 5 (5, 51), task 6 (41, 42), task 7 (44, 45), task 8 (31, 45, 46), task 9 (27, 37, 45), task 10 (6, 47, 48), and task 11 (49, 50).

Table S3. Sample size of observed dataset

Observed dataset	*n*
Total included in sample	98
Placed as guide or breeder	66
Released from program for behavioral reasons	32
Total excluded from sample	40
Transferred to external organization	1
Died	1
Released from program for medical reasons	26
Missing novel object data due to camera malfunction	1
Missing multistep problem-solving data due to failing to pass the warm-up trials	8
Missing all young adult data due to release before return to headquarters	3

Table S4. Associations between *Maternal behavior* and young adult test measures

Task no.	Task description	Estimate	German Shepherd	Labrador Retriever	Golden Retriever
1a	Isolation, anxious	0.08			
1b	Isolation, active	**0.57****			
2	Distraction	0.14			
3	Sustained attention	0.01			
4a	Memory problem solving, superior performance	0.24			
4b	Memory problem solving, accuracy	**−2.16****	0.08	0.44	**−1.72****
5	Multistep problem solving, poor performance	**0.48***			
6	Cylinder	−0.01			
7	Detour problem solving, poor performance	−0.16			
8	Greeting	0.08			
9	Ball play	0.04			
10a	Novel object, wary	**−1.85***	−0.14	**0.96****	−0.89
10b	Novel object, quiet	**−0.59***			
11a	Umbrella opening, reactivity	−0.19			
11b	Umbrella opening, recovery	**−2.03***	0.00	**1.12***	−0.92

Estimate values are listed under each breed in the event of an interaction. Estimates that were significant at $P < 0.05$ or less are bolded. Predictor variables included the following: *Maternal behavior*; breed (German Shepherd, Labrador Retriever, or Golden Retriever); litter size, 2–10; maternal parity, 1–5; sex of puppy, 1/0 (male or female); and age at return, 14–17 mo. Litter ID was entered as a random effect. $n = 98$ (32 release dogs and 66 successes). Statistical tests of significance used GEE (**$P < 0.01$; *$P < 0.05$).

Table S5. Model exploring the association between *Maternal behavior* and outcome

Predictor variables	OR	Estimate	SE	Wald	*P* value
Maternal behavior	3.39	1.22	0.34	12.98	<0.001***
Golden score	0.70	−0.35	0.33	1.11	0.292
Labrador score	0.99	−0.01	0.38	0.00	0.974
Maternal parity	1.07	0.07	0.10	0.46	0.500
Sex of puppy	0.37	−0.99	0.62	2.54	0.111
Age at return	0.71	−0.34	0.21	2.69	0.101

The dependent variable was outcome in the program, 1/0 (released from program or successfully placed as guide or breeder). Predictor variables retained were as follows: *Maternal behavior* Golden score, Golden Retriever compared with German Shepherd; Labrador score, Labrador Retriever compared with German Shepherd; maternal parity, 1–5; sex of puppy, 1/0 (male or female); and age at return, 14–17 mo. Litter ID was entered as a random effect. $n = 98$ (32 release dogs, 66 successes). Statistical tests of significance used GEE (***$P < 0.001$).

Table S6. Model exploring the association between variables comprising *Maternal behavior* and outcome

Predictor variables	OR	Estimate	SE	Wald	*P* value
Ventral nursing per puppy	4.22	1.44	0.45	10.20	0.001**
Vertical nursing per puppy	0.25	−1.39	0.24	34.57	<0.001***
Licking/grooming per puppy	1.02	0.02	0.51	0.00	0.970
Golden score	0.84	−0.17	0.47	0.13	0.716
Labrador score	1.80	0.59	0.55	1.15	0.283
Maternal parity	1.21	0.19	0.18	1.05	0.305
Sex of puppy	0.47	−0.76	0.54	2.02	0.156
Age at return	0.75	−0.29	0.22	1.80	0.180

The dependent variable was outcome in the program: 1/0 (released from program or successfully placed as guide or breeder). Predictor variables were retained as follows: ventral nursing per puppy; vertical nursing per puppy; licking/grooming per puppy; Golden score, Golden Retriever compared with German Shepherd; Labrador score, Labrador Retriever compared with German Shepherd; maternal parity, 1–5; sex of puppy, 1/0 (male or female); and age at return, 14–17 mo. Litter ID was entered as a random effect. $n = 98$ (32 release dogs, 66 successes). Statistical tests of significance used GEE (***$P < 0.001$; **$P < 0.01$).

Table S7. Using PC analysis, where applicable, to reduce variables per task in the young adult test

Task no.	Task	Measure	Type of measure	Scores into measure	Proportion variance explained, %	Fit
1a	Isolation	Anxious	PC	Time near exit (+), vocalizing (+)	54	0.77
1b	Isolation	Active	PC	Activity score (+), mobile (+)	46	0.77
2	Distraction	Distractibility	PC	Toy distraction (+), toy contact (+), food eaten (+)	61	0.77
3	Sustained attention	Attentive to human	PC	Body orient trial 1 and 2 (+), face orient trial 1 and 2 (+)	72	0.89
4a	Memory problem solving	Superior performance	PC	Solving time (–), no. correct (+), persistence (+)	69	0.97
4b	Memory problem solving	Accuracy	PC	Accuracy (+)	31	0.97
5	Multistep problem solving	Poor performance	PC	Solving time (+), perseveration (+)	80	0.89
6	Cylinder	Test trial score	Z-scored variable	Test trial score	NA	NA
7	Detour problem solving	Poor performance	PC	Test trial score (–), test trial 1 time (+), test trial 2 time (+), test trial 3 time (+)	56	0.85
8	Greeting	Willingness to interact	PC	Latency to approach (–), interact (+)	73	0.64
9	Ball play	Retrieval score	Z-scored variable	Retrieval score	NA	NA
10a	Novel object	Wary	PC	Latency to approach (+), orient (+)	58	0.69
10b	Novel object	Quiet	PC	Latency to vocalize (+)	42	0.69
11a	Umbrella opening	Reactivity	PC	Initial response (+)	35	0.64
11b	Umbrella opening	Recovery	PC	Approach (+), contact (–)	65	0.64

Table S8. Young adult testing ORs between score on each task and release from the program

		Observed dataset					Imputed plus observed dataset				
Task no.	Task description	*n*	OR	Shep	Lab	Gold	*n*	OR	Shep	Lab	Gold
1a	Isolation, anxious	109	0.87								
1b	Isolation, active	109	1.12								
2	Distraction	110	0.89								
3	Sustained attention	110	1.35								
4a	Memory problem solving, superior performance	105	1.35								
4b	Memory problem solving, accuracy	105	**2.10***				125	1.21			
5	Multistep problem solving, poor performance	100	**1.70***				120	**1.49**			
6	Cylinder	108	0.78								
7	Detour problem solving, poor performance	110	1.07								
8	Greeting	110	0.86								
9	Ball play	110	**17.12***	0.58	1.26	**21.33***	130	**1.70**	0.63	0.77	1.31
10a	Novel object, wary	109	1.35								
10b	Novel object, quiet	109	**0.60****				129	**0.61***			
11a	Umbrella-opening reactivity	110	**0.33***	0.98	**1.36**	**0.45**	130	**0.29***	0.94	**1.57***	**0.45**
11b	Umbrella-opening recovery	110	1.14								

OR values are listed under each breed in the event of an interaction. OR values that were significant at $P < 0.10$ or less are bolded. Predictor variables included the following: each task score, respectively; breed (German Shepherd, Labrador Retriever, or Golden Retriever); maternal parity, 1–5; sex of puppy, 1/0 (male or female); and age at return, 14–17 mo. Litter ID was entered as a random effect. Statistical tests of significance used GEE (**$P < 0.01$; *$P < 0.05$). The observed dataset included 98 subjects with a known outcome in the program. The imputed plus observed dataset included an additional 32 subjects that entered the training program but were subsequently released for medical reasons. In the judgment of The Seeing Eye's Director of Canine Development (who had not observed the dogs during testing and had no knowledge of their performance on tests), these subjects would have included 19 successes (59%) and 13 releases (41%) ($n = 130$; 85 successes and 45 behavioral releases). The imputed plus observed dataset was only used to aid in verifying the tasks most associated with outcome, and not in any predictive models. Shep, Shepherd; Lab, Labrador Retriever; Gold, Golden Retriever.

Table S9. Model exploring the association between young adult test performance and outcome

Predictor variables	OR	Estimate	SE	Wald	*P* value
Multistep problem solving, poor performance	1.75	0.56	0.23	5.63	0.018*
Novel object, quiet	0.43	−0.84	0.33	6.51	0.011*
Golden score	0.43	−0.85	0.31	7.64	0.006**
Labrador score	0.97	−0.03	0.41	0.01	0.941
Maternal parity	1.12	0.11	0.13	0.67	0.414
Sex of puppy	0.31	−1.18	0.59	4.03	0.045*
Age at return	0.59	−0.52	0.25	4.41	0.036*
Interaction	0.23	−1.45	0.45	10.30	0.001**
Umbrella-opening reactivity × German Shepherd	0.84	−0.18	0.46	0.16	0.691
Umbrella-opening reactivity × Labrador Retriever	1.70	0.53	0.26	4.15	0.042*
Umbrella-opening reactivity × Golden Retriever	0.40	−0.92	0.37	6.22	0.013*

The dependent variable was outcome in the program, 1/0 (released from program or successfully placed as guide or breeder). Predictor variables retained were as follows: multistep problem solving, poor performance; long latency to vocalize when presented with a novel object; an interaction between umbrella-opening reactivity and breed (German Shepherd, Labrador Retriever, and Golden Retriever); Golden score, Golden Retriever compared with German Shepherd; Labrador score, Labrador Retriever compared with German Shepherd; maternal parity, 1–5; sex of puppy, 1/0 (male or female); and age at return, 14–17 mo. Litter ID was entered as a random effect. $n = 98$ (32 release dogs, 66 successes). Statistical tests of significance used GEE (**$P < 0.01$; *$P < 0.05$).

Table S10. AUCs and 95% confidence intervals for three separate models

Predictors	AUC	95% CI
Maternal behavior PC	0.664	(0.548, 0.774)
Young adult performance	0.750	(0.640, 0.849)
Maternal behavior PC and young adult combined	0.754	(0.649, 0.851)

The 95% confidence intervals (CIs) were calculated using a bootstrapping method with 2,000 iterations.

Movie S1. Multistep problem-solving performance of a dog that was released from the program and a dog that was placed as a guide.

Movie S1

Movie S2. Latency to vocalize during novel object of a dog that was released from the program and a dog that was placed as a guide.

Movie S2