Appendix S1: Additional information and analyses.

S1.1: Original Eurasian otter monitoring study

Monitoring took place in the Upper Lusatian Heath and Pond Landscape in eastern Saxony, Germany (51°20'N, 14°19'E). The aim of the original monitoring study was to collect scats of the Eurasian otter (*Lutra lutra*) for population and spatial analyses. However, scat of other species, mainly the American mink (*Neovison vison*), resampled otter scats in size and shape and did not show the typical smell or contents of mink scats. Therefore, collected scats finally include that of Eurasian otter and American mink visually identified as otter scat.

Sampling was conducted on five consecutive days each spring between 2006 and 2012 (except 2009). The study area comprised 64 ponds clustered in 7 pond areas, each composed of 8 to 13 ponds of varying size (0.36-39.6 ha) connected by ditches and streams (Lampa et al. 2015; Fig. 1). Sampling followed a clearance sampling design as described in Klenke et al. (2008) and Lampa et al. (2015). For each sample, we recorded whether the marking site was a hotspot (> 5 fresh scats), the surface (actively exposed (e.g. scratch piles), passively exposed (e.g. stones, roots, sticks, grass tussock), on human structures, in front of a den, or non-exposed), the amount (few (< 1.5 cm), moderate (1-3 cm), or much (> 3 cm)), the sliminess (spraint, spraint plus mucus, jelly), and the color if it was jelly. Samples were genetically assigned to species (otter, mink, other) if DNA extraction was successful (Lampa et al. 2015).

This study area was surveyed again for testing relative and absolute detection abilities of scat detection dogs versus human searches in 2017. The field study is described in the main text. Figure 1 highlights the four pond areas used in the field study. It also displays the starting point and direction of the four transects used for comparisons.

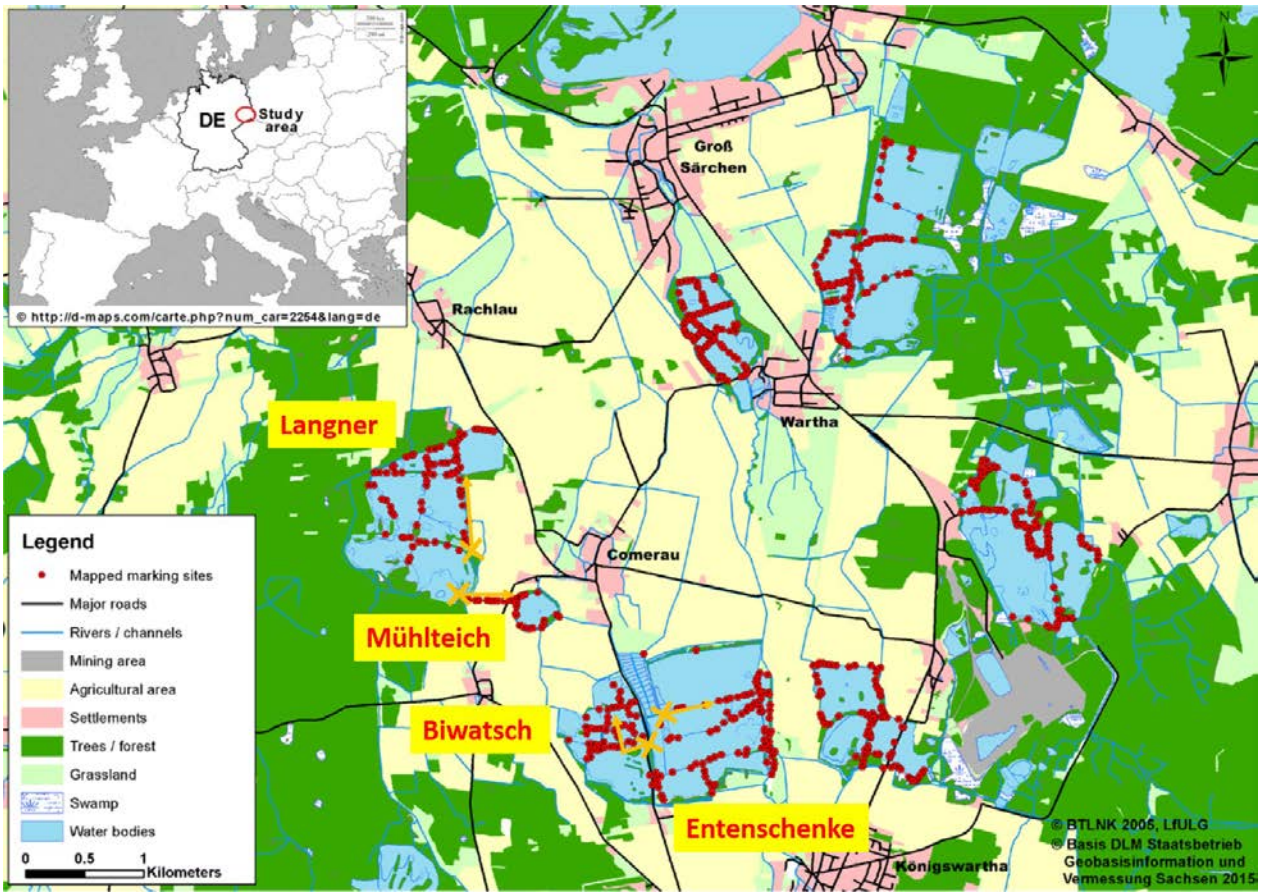

Figure S1: Study area map with recorded otter marking sites. Location of otter marking sites (red hexagons) in seven pond areas and one single pond in the Upper Lusatia (Saxony, Germany), where we searched for fresh faeces for genetic capture-mark-recapture (CMR) analyses (2006–2012). Main land use types of the surrounding area are outlined. Names of the ponds monitored for transect comparisons, their starting points (orange crosses) and the directions (orange arrows) of the four transects are highlighted. Map modified after Lampa et al. (2015).

S.1.2 Detection dogs pre-test and discrimination training

To examine whether a dog had understood the procedures on the scent box, we performed a pre-test. Therefore, we placed the scat of one target individual (otter for otter scat detection dogs, mink for mink scat detection dogs) in one cup using cotton buds and a clean cotton bud only in each of the other four cups (blanks) to avoid the dogs alerting at plastic or cotton later on. Throughout ten trials, we randomly varied the location of the target among the five cups and used ten different individual scat samples. We asked the dogs to alert at the target in all ten trials. All pre- tests were conducted double-blind, i.e. an assistant hid the samples while the handler and the dog were waiting in another room. The assistant then left the room while the handler and the dog entered. They conducted the trial on the scent box alone. The handler would say the number of the cup indicated by the dog and the assistant, still being in another

room, would approve if the choice was correct. All four dogs passed the pre-test to detect 100% of the target species scats in their first attempt.

When a dog passed the pre-test, we started the discrimination training. To introduce the non-target scent, i.e. the mink for otter scat detection dogs and the otter for mink scat detection dogs, we used the method of errorless discrimination (Gadbois and Reeve 2014). We started with a huge amount of mixed target scats and a very small amount of mixed non-target scats making the target scent more attractive. Over the course of the training, we decreased the amount of the target scat and increased the amount of the non-target scat until the amount of both did not influence the correct alert at the target anymore. This includes that even if the amount of non-target scat was substantially higher than that of the target scat, the dog would clearly indicate the target scat. During training, we also moved from mixed samples to single individual samples to obtain generalization throughout individuals of the target species but discrimination against any other individual of the non-target species.

Discrimination training was adapted to each dog separately, and thus the number of trials per training session and the total number of training sessions differed between them. Specifically, the number of training sessions varied among dogs from 7 to 16 (Table S1). From the first presentation of a mink sample to the last test trial, the otter scat detection dogs Bagheera and Zammy correctly indicated 94% and 89% of all otter scats and ignored 94% and 89% of all mink scats, respectively. Thus, wrong alerts included only the non-target species but no blanks. In comparison, the mink scat detection dogs Cue and Zoey correctly indicated 79% and 89% of all mink samples and ignored 84% and 90% of all otters, respectively (Table S1). The difference between true-positive and true-negative responses emerged from the fact that both dogs alerted at 21 and 7 blanks, respectively. In almost all of those cases, they pointed to the location where the scat had been hidden in the previous trial. During training they started to correct themselves immediately afterwards but we still denoted if the initial alert was at a blank.

Table S1: Number of training sessions and trials per dog and the total amount of dog alerts during discrimination training. The number of trials includes all trials from first presentation of the non-target scat to the final test including all repetitions if a trial was wrong in the first attempt.

Dog	# sessions	# trials	true-positive	false-positive	true-negative	false-negative
Bagheera	7	134	126 (94%)	8 (6%)	126 (94%)	8 (6%)
Zammy	7	80	71 (89%)	9 (11%)	71 (89%)	9 (11%)
Cue	16	394	313 (79%)	81 (21%)	329 (84%)	65 (16%)
Zoey	14	321	285 (89%)	36 (11%)	289 (90%)	32 (10%)

S1.3: Do humans assign species based on color?

Since humans assigned the species mainly visually, we further analyzed whether test candidates in the species assignment tests used the scat color for species assignment. We applied Fisher's exact rank test for each experience level separately.

All but two persons assigned scats of one or two colors to one species, mostly all green or dark-green samples being otter or all black samples being either otter or mink. This separation by color was almost statistically significant for experienced persons (Fisher-Test, $p = 0.07$) and experts (Fisher-Test, $p < 0.001$), but not for beginners (Fisher-Test, $p = 0.24$) despite being significant for two of the three persons if considered separately (Fisher-Test, $p < 0.05$). We presume the reason for this is that one of these two persons assigned all dark scats to otter, while the other did the opposite. The third person was not influenced by scat color.

S1.4: Evaluation of dog performance in the field

During otter monitoring in the Upper Lusatia, each dog team additionally noted the dog's alert (strong, light, aggressive, avoidance) and behavior (cooperative, unresponsive, frustrated, uninterested, fatigued, distracted) during both transect comparisons and otter presence monitoring (MacKay et al. 2008). Dogs alerts were mostly strong (Bagheera: 67.1%, Zammy: 58.2%) or moderate to light (Bagheera: 25.4%, Zammy: 37.3%). Remaining alerts were aggressive alerts through urine markings or scratch marks. Dogs were not rewarded for such aggressive alerts. The experienced detection dog Bagheera was behaving cooperatively in all cases, while the recruit Zammy was behaving cooperatively in 82%, distracted in 9%

and fatigued in 9%. These findings highlight that even under pressure and during a five-day monitoring, the dogs showed high working abilities to their jobs and no change in their performance.

S1.5. Species assignments for humans and dogs

Table S2: Averaged number of assignments (in %) of the species assignment test of humans grouped by their level of experience, as well as the overall average assignment. Rows show the true species, columns show the decision by the test group.

pointed / **true**	**Eurasian otter**			**American mink**		
Eurasian otter	Beginner 58%	Experienced 76%	Expert 91%	Beginner 42%	Experienced 24%	Expert 9%
	Average 76%			Average 24%		
American mink	Beginner 41 %	Experienced 33%	Expert 14%	Beginner 59%	Experienced 67%	Expert 86%
	Average 28%			Average 72%		

Table S3: Species matching during the discrimination test (in %) separated by dog, with the target species per dog being highlighted in yellow. Rows show the true species, columns show the decision by the dog.

pointed / **true**	**Eurasian otter**		**American mink**	
Eurasian otter	Bagheera 100%	Zammy 100%	Bagheera 0%	Zammy 0%
	Cue 80%	Zoey 85%	Cue 20%	Zoey 15%
American mink	Bagheera 0%	Zammy 6%	Bagheera 100%	Zammy 94%
	Cue 0%	Zoey 0%	Cue 100%	Zoey 100%

References

Gadbois, S., and C. Reeve. 2014. Canine olfaction: scent, sign, and situation. Pages 3–29 *in* A. Horowitz, editor. Domestic dog cognition and behavior: the scientific study of Canis familiaris. Springer, Berlin-Heidelberg, Germany.

Klenke, R., A. Tschierschke, B. Gruber, U. L. F. Hempel, S. Geissler, C. Helm, C. Hofmann, J. Kalbitz, J. Kaulfuss, Z. Liu, and K. Henle. 2008. Räumliche und zeitliche Aktivitätsmuster von Fischottern (*Lutra lutra* Linnaeus, 1758) anhand von frischen Markierungsfunden in einer Teichlandschafe. Säugetierkundliche Informationen 6:209–226.

Lampa, S., J. B. Mihoub, B. Gruber, R. Klenke, and K. Henle. 2015. Non-invasive genetic mark-recapture as a means to study population sizes and marking behaviour of the elusive Eurasian otter (*Lutra lutra*). PLoS ONE 10:e0125684.

MacKay, P., D. A. Smith, R. A. Long, and M. Parker. 2008. Scat detection dogs. Pages 183–222 *in* R. A. Long, P. MacKay, W. J. Zielinski, and J. C. Ray, editors. Noninvasive survey methods for carnivores. Island Press, Washington-Covelo-London.