

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

Innovative problem-solving in wild hyenas is reliable across time and contexts

- 1. Lily Johnson-Ulrich^{1,2*}
- 2. Kay E Holekamp^{1,2}
- 3. David Z Hambrick³

¹Department of Integrative Biology, Michigan State University, East Lansing, MI, 48824, USA
²Ecology, Evolutionary Biology, & Behavior Program, Michigan State University, East Lansing, MI, 48824, USA
³Department of Psychology, Michigan State University, East Lansing, MI, 48824, USA
*john3923@msu.edu

Corresponding author:
Lily Johnson-Ulrich
288 Farm Lane, Rm 203
Natural Sciences Bldg
East Lansing, MI 48823
U.S.A.
+1 301 676 4128
john3923@msu.edu

26 **Abstract**

27 Individual differences in behavior are the raw material upon which natural selection acts, but
28 despite increasing recognition of the value of considering individual differences in the behavior
29 of wild animals to test evolutionary hypotheses, this approach has only recently become popular
30 for testing cognitive abilities. In order for the intraspecific approach with wild animals to be
31 useful for testing evolutionary hypotheses about cognition, researchers must provide evidence
32 that measures of cognitive ability obtained from wild subjects reflect stable, general traits. Here,
33 we used a multi-access box paradigm to investigate the intra-individual reliability of innovative
34 problem-solving ability across time and contexts in wild spotted hyenas (*Crocuta crocuta*). We
35 also asked whether estimates of reliability were affected by factors such as age-sex class, the
36 length of the interval between tests, or the number of times subjects were tested. We found
37 significant contextual and temporal reliability for problem-solving. However, problem-solving
38 was not reliable for adult subjects, when trials were separated by more than 17 days, or when
39 fewer than seven trials were conducted per subject. In general, the estimates of reliability for
40 problem-solving were comparable to estimates from the literature for other animal behaviors,
41 which suggests that problem-solving is a stable, general trait in wild spotted hyenas.

42 **Introduction**

43 The questions of how and why cognition evolves across the animal kingdom remain unresolved
44 despite more than a century of intensive research. The most common approach to addressing
45 these questions has been to compare average levels of cognitive performance among species [1–
46 5]. In this *interspecific* approach, individual differences within species are treated as random
47 error (or “noise”). Recently, there has been growing recognition of the value of using individual
48 differences to test evolutionary hypotheses—the *intraspecific* approach [6]. Intraspecific studies
49 of free-ranging populations are especially valuable for understanding cognitive evolution,
50 because individual variation is the raw material on which natural selection acts. This approach
51 allows researchers to examine the causes of cognitive variation in an ecologically valid context
52 and also to examine the fitness consequences of this variation [7–9]. Despite this recognition, in
53 the field of cognitive ecology there have been few attempts to empirically test the hypothesis
54 that measures of cognition reflect stable, general traits, meaning traits expected to influence
55 performance across time and across a wide range of situations [10,11].

56 The hypothesis that a cognitive measure reflects a stable, general trait predicts that the
57 measure should have a high degree of reliability: an animal that receives a high score on the
58 measure at one point in time and in one context should receive a high score at later points in
59 times and in other contexts, and the performance of animals receiving low scores on the
60 measure should be similarly consistent. As a psychometric concept, *reliability* refers to the
61 amount of error contained in a measure, as reflected in the stability of the measure across
62 contexts and time. Although reliability is synonymous with the term ‘repeatability’, which
63 refers to consistent individual differences in the behavior of non-human animals [12], we use
64 reliability because it is a well-defined psychometric term used in diverse literatures on
65 individual differences in psychological traits, including cognitive abilities in both humans and
66 non-human animals. It is especially important to demonstrate reliability of measures reflecting
67 animal cognition in the wild, because there are many potential sources of error, including both
68 external factors (e.g., weather, presence of conspecifics) and internal factors (e.g., hunger,
69 stress) [9,13,14].

70 In this study, we assessed intra-individual reliability of *innovativeness* across time and
71 context in wild spotted hyenas using a problem-solving paradigm. Defined as the ability to solve

72 a novel problem or use a novel behavior to solve a familiar problem, innovativeness is among
73 the most commonly measured cognitive abilities in non-human animals [15]. Although there has
74 been a great deal of interest in the relationship between innovativeness and variables such as
75 brain size, ability to invade new habitat, and life history traits in a diverse array of taxa, formal
76 attempts to evaluate the reliability of problem-solving paradigms used to measure
77 innovativeness remain very rare. In a meta-analysis, Cauchoix et al. [11] identified only six
78 publications reporting reliability for any measure of cognitive performance; of these only two
79 measured innovative problem-solving in wild subjects, and both were in birds. Thus, there is a
80 pressing need to examine the reliability of innovative problem-solving in other wild animals,
81 especially in wild mammals. Furthermore, most studies only measure cognition at two time
82 points and across two to four different tasks [16–19], and there has been very little research
83 examining how reliability might vary based on the number of measures or the length of the
84 interval between measures, nor how reliability might vary within species among different age-
85 sex classes [e.g. 13]. Our ignorance here is due, in part, to the numerous logistical challenges of
86 experimentally measuring innovativeness repeated times in the same subjects—a problem that is
87 particularly pronounced in wild subjects where locating and enticing individuals to perform
88 cognitive tests even once can be difficult, and tracking individuals for repeated testing may be
89 impossible in many species. However, for the intraspecific approach in wild animals to be
90 useful for testing evolutionary hypotheses, researchers *must* provide evidence that measures of
91 innovative-problem solving reflect stable traits, and that estimates of reliability are robust
92 against numerous sources of variation in testing environment and methodology.

93 The spotted hyena is well-established as a model organism for testing hypotheses about
94 the evolution of cognition [20] and innovativeness has previously been measured in both captive
95 and wild hyenas [21–23], but the problem-solving paradigms used to measure innovativeness
96 were never previously tested for reliability except by Johnson-Ulrich et al. [24]. Here we
97 measured reliability (R) by calculating intraclass correlation coefficients (ICC), which are
98 commonly used in behavioral ecology to assess the reliability of behavioral traits within
99 individuals [25]. An ICC estimates the amount of variation in the response variable explained
100 by random effects or grouping factors in mixed hierarchical models. Ultimately, we found
101 significant reliability for problem-solving performance in wild spotted hyenas and demonstrate
102 how estimates of reliability vary across tasks, trials, age-sex classes, the temporal interval
103 between observations, and the total number of observations.

104 **Results**

105 Seventy-two hyenas participated in 694 test trials with a multi-access box (MAB). MABs are
106 problem-solving paradigms used to measure innovativeness. The MAB used in the present study
107 was a metal box with one of four different doors on each vertical face (Fig. 1). Each door
108 required a unique motor behavior to open, but all four doors opened to the same common
109 interior from which a hyena could retrieve bait. ~~We defined problem-solving as the ability to~~
110 ~~successfully open a door to the MAB.~~ Hyenas were given repeated trials, and after a hyena
111 opened the same door on three of four consecutive test trials, that door was blocked, forcing the
112 hyena to open a different door to retrieve the bait (Fig. 2). Testing was thus divided into four
113 ‘phases’ in which hyenas were required to use four different doors to open the MAB
114 (Supplementary Fig. S1). Overall, our sample included an adequate representation of each age
115 and sex class with 17 adult females, seven adult males, 13 subadult females, 17 subadult males,
116 10 female cubs, and nine male cubs. Out of these 72 hyenas, 23 opened the MAB at least once
117 (mean = 2.74 doors, SD = 1.39) and 11 opened each of the four doors to the MAB at least once

118 across their trials. Because we collected data with more hyenas that never opened the MAB
119 (N=49 hyenas, n=376 trials) than data with hyenas who opened the MAB at least once (N=23
120 hyenas, n=318 trials), and because including the former hyenas would lead to zero-inflation and
121 an inflated reliability estimate, we excluded data from hyenas that never opened the MAB from
122 our analyses. Instead, we only assessed the reliability of problem-solving for hyenas that opened
123 the MAB at least once. Among these 23 hyenas our sample included five adult females, two
124 adult males, five subadult females, six subadult males, one female cub, and four male cubs.

125 **Contextual reliability of problem-solving ability across different doors**

126 Contextual reliability is typically assessed by comparing performance across different tasks that
127 measure the same cognitive ability [11,26]. Because each of the four doors to the MAB required
128 a different motor behavior, we first investigated the contextual reliability of problem-solving
129 with a model examining the likelihood of solving each of the four different MAB solutions
130 (door knob, slot, push, drawer) at least once. Because innovation is defined as using a *novel*
131 behavior to solve a problem [15], hyenas only ‘innovated’ when they solved each door of the
132 MAB for the very first time. Thus, our estimate of reliability for problem-solving across doors
133 provides a valuable indication of the reliability of ‘innovativeness’, in the strictest sense, among
134 hyenas. In this model our response variable was a binary variable indicating whether or not a
135 hyena had solved each of the four doors to the MAB at least once (Table 1, Model 1). Each
136 hyena received four dichotomous scores for each of the four unique MAB doors, with a score of
137 one indicating that they solved a door at least once and a score of zero indicating that they never
138 solved that particular door despite contacting the MAB on multiple trials. We included age class
139 as a fixed effect in this model (see Methods: Statistical Analysis); hyenas in the cub (GLMM:
140 Odds ratio = 0.05, P = 0.058) and subadult (GLMM: Odds ratio = 0.07; P = 0.055) age classes
141 were less likely than adults to solve each door. Reliability was determined by calculating
142 adjusted ICCs with the R package rptR [27]. Adjusted ICC values are calculated as a ratio of the
143 random effects variance to the combined random effects and residual variance. Of the 23 hyenas
144 that opened the MAB at least once, we found that problem-solving performance across doors
145 was moderately but significantly reliable (Likelihood ratio test: R = 0.40, P = 0.001; Table 1:
146 Model 1). Thus, problem-solving ability was significantly reliable when assessed in the variable
147 contexts of the MAB’s four doors, each of which required a unique solution.

148 **Temporal reliability of problem-solving across trials**

149 In addition to contextual reliability, the temporal reliability of cognitive traits is commonly
150 assessed by comparing performance across repeated trials with the same cognitive test [e.g.
151 problem-solving: 11,17,19,28]. Because problem-solving performance was moderately reliable
152 across different doors, we next examined the reliability of problem-solving performance across
153 each subject’s trials, regardless of the specific doors used, in order to investigate temporal
154 reliability. We gave each hyena multiple trials with the MAB in order to give subjects the
155 opportunity to solve its different doors and examine performance across different phases of
156 testing. Although hyenas are not strictly ‘innovating’ when they open a MAB door that they’ve
157 previously solved, most studies on innovative problem-solving conduct repeated trials to
158 compare the acquisition of innovations across individuals or assess their spread through
159 populations [e.g. 21,28–32], so investigating the reliability of problem-solving performance
160 across trials is relevant for future research.

161 In this model, and all subsequent models, our response variable was a binary variable
162 indicating whether a hyena opened or failed to open a door of the MAB, irrespective of which

163 specific door it was working on. On average hyenas were successful in 54.5% of trials (SD =
164 27.0%, N = 23 hyenas, n = 318 trials). Because temporal reliability may be influenced by
165 learning and experience [11] we included a fixed effect of trial number in order to control for
166 the number of previous trials in which each hyena participated. We also included age class and
167 phase number as fixed effects (see Methods: Statistical Analysis). We found that cubs (GLMM:
168 Odds ratio = 0.31, P = 0.096) and subadults (GLMM: Odds ratio = 0.32, P = 0.046) were less
169 likely than adults to have successful trials with the MAB. Hyenas were also less likely to solve
170 the MAB at later than earlier phases of testing (GLMM: Odds ratio = 0.54, P = 0.044), which
171 probably represents the increasing difficulty across phases. Trial number had a significant
172 positive effect on the odds of a trial being successful (GLMM: Odds ratio = 1.11, P = 0.050),
173 which suggests that prior experience or learning with the MAB was important, but the effect of
174 trial number on the odds of success was relatively small compared to the effects of age class and
175 phase. Furthermore, these fixed effects only explained half as much variation in the response
176 variable ($\text{Var}_F = 0.08$; Table 1: Model 2:) as that explained by hyena ID ($\text{Var}_R = 0.16$; Table 1:
177 Model 2). Among the 23 hyenas that opened the MAB at least once, problem-solving
178 performance was significantly reliable (Likelihood ratio test: $R = 0.18$, $P < 0.001$; Table 1:
179 Model 2). This result suggests that some hyenas were consistently more likely to open the MAB
180 while others were consistently less likely to open the MAB, even after controlling for the
181 number of previous trials, the phase of testing, and the hyena's age class.

182 **Reliability of innovative problem-solving within different age-sex classes**

183 Next, we inquired whether temporal reliability varied among individuals in different age-sex
184 classes. For example, some evidence suggests that female animals exhibit more reliable
185 behavior than males [12]. Furthermore, it seems reasonable to expect that juveniles, who are
186 still developing, might exhibit behavior that is less reliable than that of adults in addition to
187 showing slightly worse performance with the MAB than adults. To compare reliability within
188 different age and sex classes we partitioned our dataset into five different categories: females,
189 males, adults, subadults, and cubs in order to create five different models examining reliability
190 for each age-sex class independently. We included age class as a fixed effect in the female-only
191 and male-only models, and we also included both trial number and phase of testing as fixed
192 effects in all models (see Methods: Statistical Analysis). We did not include sex as a fixed effect
193 in each age class model because problem-solving did not vary with sex (Supplementary Tables
194 2-3). We found that most age and sex classes showed moderate levels of reliability (Likelihood
195 ratio test: $R = 0.21 - 0.33$, $P < 0.001$; Table 1: Models 3.1-3.5; Fig. 3) with the exception of
196 adult hyenas, for which the reliability of problem-solving was not significantly different than
197 zero (Likelihood ratio test: $R = 0.07$, $P = 0.11$; Table 1: Model 3.3).

198 **Reliability of innovative problem-solving across different timespans**

199 Although most test trials within subjects (49.37%) were conducted less than 1 day apart, the
200 average number of days between testing sessions was 19.41 ± 56.00 days (median = 0, range =
201 0 – 301 days). We were interested in whether temporal reliability between any given trial and
202 the trial that followed it was affected by the amount of time between trials. To do this, we
203 created a dataset where we paired each subject's trial with the trial that followed it and
204 calculated the number of days elapsing between the two trials. We next partitioned this dataset
205 into trials that occurred less than one day apart, one to three days apart, four to sixteen days
206 apart, and more than 17 days apart. The number of bins and the date range included in each bin
207 were chosen to distribute the number of trials across each date range as equally as possible. We

208 then calculated reliability between pairs of trials for each of these datasets (Table 1: Models 4.1-
209 4.4; Fig. 4). We included age class, phase of testing, and trial number in these models (see
210 Methods: Statistical Analysis). We found that reliability was extremely high for trials collected
211 on the same day (Likelihood ratio test: $R = 0.93$, $P < 0.001$; Table 3: Model 9.1), but reliability
212 became non-significant when trials were separated by 17 or more days (Likelihood ratio test: R
213 $= 0.10$, $P = 0.235$; Table 1: Model 4.4).

214 **Reliability of innovative problem-solving across different numbers of trials**

215 Finally, we were interested in how the varying number of trials collected per hyena might affect
216 estimates of temporal reliability. On average, hyenas received 13.8 ± 7.3 trials (median = 15
217 trials, range = 2 – 26 trials). Although we found modest levels of temporal reliability, we were
218 interested in how our estimates might have changed if we'd only sampled hyenas a set number
219 of times. Collecting a greater number of trials per hyena could, in theory, increase the accuracy
220 of estimates about their problem-solving ability and therefore increase reliability; however,
221 increasing the number of trials can also strengthen learning and memory, which may ultimately
222 reduce estimates of reliability if all individuals eventually converge at a high level of
223 performance [11]. On the other hand, because we were testing hyenas in the wild, larger number
224 of trials were also more likely to take place across different testing sessions, different timespans,
225 and under variable environmental conditions which could, in theory, decrease estimates of
226 reliability due to increased variability with increasing numbers of trials. To estimate reliability
227 for varying numbers of trials, we calculated reliability for hyenas in nine models where we
228 included only their first two to ten trials. We found that estimates for the reliability of problem-
229 solving performance were not significant until we had sampled each hyena seven times
230 (Likelihood ratio test: $R = 0.13$, $P = 0.026$, Table 1: Model 5.6; Fig. 5). With seven or more
231 trials estimates of reliability were modest, but nonetheless significantly greater than zero
232 ($R=0.13-0.20$; Table 1: Models 5.6-5.9).

233 **Discussion**

234 Overall, our results suggest that innovative problem-solving ability is a stable, general trait in
235 wild spotted hyenas. Our estimates for the reliability of problem-solving performance are
236 comparable to the average reliability of other behaviors in wild animals [12], and also to the
237 average reliability of other cognitive measures in both captive and wild animals [11]. However,
238 building on previous findings, we further present evidence that, with a few important
239 exceptions, problem-solving performance is reliable across context, time, age-sex class, the
240 interval between observations, and the number of observations.

241 We found moderate levels of reliability for problem-solving performance across the four
242 different MAB doors. These doors represent four different motor tasks, each designed to
243 measure innovativeness, and we found that hyenas who innovated with one door to the MAB
244 were moderately likely to innovate with the other three doors to the MAB (Table 1: Model 1).
245 This result is similar to studies in wild and captive birds that have generally found consistent
246 performance among problem-solving tasks requiring different motor actions [17,18,34,35].

247 Next, we also evaluated the temporal reliability of problem-solving performance across
248 all trials, irrespective of the specific door used to open the MAB. We found a modest, but
249 significant, level of reliability for problem-solving performance across trials (Table 1: Model 2).
250 Because trials were conducted across a wide variety of socio-ecological conditions we were
251 impressed to find hyenas demonstrate even this level of consistency in performance. Trial
252 number did have a significant effect in this model, which suggests that learning may have

253 played a role in shaping consistency across trials (Fig. 6); however, the amount of variation
254 explained by subject ID in these models was twice that explained by the fixed effects, which
255 included trial number. This result is also consistent with other cognitive studies; a meta-review
256 of the reliability of cognitive abilities similarly found that repetition number usually had an
257 important effect on cognitive performance [11]. However, this meta-review also found that
258 adjusting estimates of R for repetition number usually did not increase R and, so the authors
259 concluded that repetition numbers largely had negligible effects on estimates of temporal
260 reliability. Likewise, in most of our models our adjusted R values were only modestly larger
261 than the total amount of variation explained by the random effects. While most studies of
262 problem-solving performance do provide evidence that subjects improve their performance over
263 trials, this improvement is typically gradual, which suggests that subjects do not perfectly
264 remember the motor behaviors used to innovate during their first trial [14,21,23,29,33]. Instead,
265 the literature suggests that behaviors such as motor diversity or flexibility may be key for
266 successful problem-solving and that these behaviors, even though they might interact with
267 memory, are independent from learning [33,36–38]. Ultimately, a great deal of variation in
268 problem-solving performance was left unexplained by our models, an unsurprising result given
269 that our subjects were wild, free-ranging hyenas tested in an uncontrolled environment. Future
270 research investigating this remaining variation may shed light on the various individual
271 behaviors or socio-ecological conditions that favor successful problem-solving.

272 Next, we examined the reliability of problem-solving performance within different age-
273 sex classes. Both female and male hyenas showed similar levels of reliability for problem-
274 solving performance (Table 1: Models 3.2.-3.3) with a slight trend towards higher reliability in
275 females. These results are similar to results for behavior across animals more generally; a meta-
276 review of the reliability of animal behavior found that females tend to show slightly more
277 reliable behavior than males when mate-preference behavior is excluded [12]. When we
278 compared the reliability of problem-solving performance across hyena age classes, we found
279 that subadults and cubs showed slightly higher estimates of reliability than did adults, which is
280 the opposite of what we'd expected, especially because subadults and cubs were significantly
281 less likely to solve the MAB. A meta-review of the reliability of animal behavior found that
282 adults and juveniles tend show similar levels of reliability across behaviors [12]. In wild hyenas,
283 it may be that adults must contend with a wider variety of distractions than non-reproductively
284 active individuals that are still largely reliant on maternal support for survival [39]. However, it
285 may also be that higher reliability among cubs and subadults compared to adults is directly
286 related to their poorer performance with the MAB compared to adults. Cubs and subadults were
287 successful on $45.8\% \pm 32.3\%$ and $47.4\% \pm 27.6\%$ of trials respectively whereas adults were
288 successful on $72.1\% \pm 13.7\%$ of trials. In adults, lower reliability here could be a result of a
289 ceiling effect where the relatively high success rate and lower variability across trials in adults
290 reduces the amount of variation explained by individual differences.

291 In general, estimates of reliability are higher for behavioral observations that are made
292 closer together in time [12]. Here, we found remarkably high reliability for problem-solving
293 performance within pairs of trials separated by less than a day (Table 1: Model 4.1). We also
294 found low to moderate reliability for trials separated by as much as 16 days (Table 1: Model
295 4.2-4.3). Only when trials were separated by more than 17 days did we find no significant
296 reliability within pairs of trials. The lack of reliability among pairs of trials separated by 17 days
297 or more could reflect a limitation of hyenas' long-term memory, but research with wild spotted
298 hyenas suggests that they are able to efficiently open a previously solved puzzle box even after

299 delays of over a year (unpublished data). In addition to memory, both internal and external
300 environmental conditions (e.g. hunger, social environment) are also much more likely to vary
301 across larger than shorter time spans. That hyenas still show some consistency even with as
302 much as two weeks separating trials is important because it can be extremely difficult to
303 consistently locate subjects for repeated testing, especially in animals like spotted hyenas that
304 live in fission-fusion societies occupying large territories.

305 In a meta-review of reliability in earlier animal behavior research, reliability estimates
306 were generally not affected by the number of observations per individual [12]. Here, we found
307 low to no reliability for problem-solving performance when fewer than seven trials were
308 conducted per individual. Part of this result may be due to sample size, with just twenty-three
309 hyenas that solved the MAB at least once, we were only able to include 46 trials in Model 5.1
310 (Table 1). However, part of this may also reflect high intra-individual variability in problem-
311 solving performance for subjects in their first several trials. Although most hyenas opened the
312 MAB on their first trial (median = 1 trial, mean \pm SD = 1.96 ± 1.26 trials), the highest trial
313 number in which any of these subjects opened the MAB for the very first time was the fourth
314 trial. No subjects ever solved the MAB after four consecutive failures, despite having
315 subsequent opportunities to do so. For this reason we used a conservative criterion of at least
316 five consecutive failures to classify hyenas as non-innovative (N=49 hyenas, n=376 trials),
317 though their trials were not included in our models examining reliability. The lack of reliability
318 across our subjects' early trials differs from the results obtained from a meta-review of
319 reliability of animal behavior generally and probably reflects the difficulty of getting accurate
320 measures of animal cognition, especially in wild subjects, where many other internal and
321 external factors may affect the way a subject interacts with a test apparatus, independent of its
322 actual cognitive abilities. Our results suggest that, if researchers are testing problem-solving in
323 wild subjects, they should aim to collect many trials per subject to ensure accurate estimates of
324 their problem-solving ability, and aim to identify a minimum number of trials per subject for
325 inclusion in analyses. In hyenas, it appears that 5-7 trials per subject may be required to
326 adequately measure their problem-solving ability. In total, we deployed the MAB an average of
327 88.5 ± 34.72 (N = 4 clans) times in each of four study groups in order to identify initial
328 successful trials for all 23 innovative hyenas (see Methods: Testing Protocol).

329 Our study offers an important demonstration of the reliability of innovative problem-
330 solving in a wild mammal. However, reliability does not necessarily correlate with validity.
331 Previous research has heavily debated the conceptual validity of problem-solving paradigms for
332 measuring innovativeness [14,36,40,41]. Although this debate is not entirely settled, researchers
333 have found that the behaviors leading to spontaneous innovations in the wild are very similar to
334 the behaviors that underlie experimentally-observed innovations using problem-solving
335 paradigms [36], which strongly suggests that problem-solving paradigms are valid for
336 measuring innovativeness. However, it is also important to consider the ecological validity of a
337 paradigm and tasks should be designed with a species' underlying capabilities in mind. We
338 designed a multi-access box that required spotted hyenas to use behaviors that are part of their
339 natural foraging repertoire to solve a novel problem for a food reward. This kind of puzzle box
340 is sometimes called a novel extractive foraging puzzle because it requires subjects to extract
341 food from a container. Spotted hyenas are dietary generalists and mammalian bones, which
342 represent an important part of their diets, require a moderate degree of extractive foraging to
343 access the marrow within. Therefore, it is not surprising that spotted hyenas were able to
344 innovate with this kind of problem-solving paradigm. However, for animals that never use

345 extractive foraging in the wild, problem-solving paradigms like the one used in the current study
346 might not be ecologically valid tools for assessing innovativeness.

347 In conclusion, it appears that, even with the many challenges posed by testing animals in
348 the wild, we were nevertheless able to reliably measure innovative problem-solving ability in
349 hyenas. Overall, our results on reliability complement the literature on the validity of innovative
350 problem-solving paradigms, and we conclude that innovative problem-solving paradigms are
351 reliable tools for measuring individual variation in cognitive performance.

352 **Methods**

353 **Study site and subjects**

354 We tested innovativeness in four neighboring spotted hyena clans within the Maasai Mara National Reserve, Kenya
355 between June 2016 and November 2017. These clans ranged in size from 30 to 55 adult hyenas. Spotted hyena
356 clans represent distinct social groups that are made up of multiple unrelated females, their offspring, and adult
357 immigrant males. Clans are structured by strict linear dominance hierarchies, with an alpha female and her
358 offspring at the top, followed by lower-ranking females and their offspring, with adult immigrant males occupying
359 the lowest rank positions. Births occur year-round and unrelated females raise their offspring together at a
360 communal den. Female hyenas stay in their natal clan throughout their lives, whereas male hyena usually disperse
361 to join new clans when they are 24-60 months old, after they reach sexual maturity [42,43].

362 All subjects were identified by their unique spot patterns and ear damage. Hyenas of all age classes and
363 both sexes were included in the study. All subjects were sexed within the first few months of life based on the
364 genital morphology [39]. Age classes were based on life history stage [44]. Cubs were defined as hyenas that were
365 still dependent on the communal den for shelter; on average, Mara cubs become den-independent around 9-12
366 months of age [44]. Subadults were hyenas who were den-independent but had not yet reached sexual maturity.
367 Adults were hyenas that had reached sexual maturity. In females, sexual maturity was determined by the
368 observation of mating, visual evidence of first parturition, or the female reaching three years of age, whichever
369 came first [45]. In males, sexual maturity was determined by dispersal status, males who were still present in their
370 natal clan at testing were classified as subadults and immigrant males were classified as adults.

371 **Multi-access box paradigm for measuring repeated innovation**

372 We tested innovativeness in wild spotted hyenas using a multi-access box designed for use with mammalian
373 carnivores [23]. The multi-access box (hereafter, ‘the MAB’) is a problem-solving paradigm, also known as an
374 artificial or novel extractive foraging task, where subjects must solve a novel problem to obtain a food reward. In
375 contrast to traditional problem-solving tasks, MAB paradigms typically offer multiple solutions to the same puzzle,
376 each requiring its own unique behavior pattern. As a condensed battery of tasks, the MAB paradigm allows
377 researchers to measure innovation, not just once, but multiple times across different doors [46].[46]. We chose to
378 use a MAB paradigm because it allowed us to compare reliability across repeated trials within the same door to
379 reliability across different doors. Reliable success with the same door across trials may be a result of individual
380 learning rather than a result of a stable cognitive trait. However, if individuals reliably innovate by opening
381 multiple unique doors to the MAB this would suggest that innovativeness is a stable cognitive trait. The MAB in
382 the current study was a steel box, measuring 40.64 x 40.64 x 40.64 cm (length x width x height), with four unique
383 doors, each requiring a different motor behavior, that could be used to access a common interior baited with a food
384 reward (Fig. 1). We used this MAB previously to test repeated innovation in captive hyenas; for more information
385 about the design specifications see Johnson-Ulrich et al. [23].

386 **Test protocol**

387 We conducted all testing between 0630 to 1000 hours and 1700 to 1830 hours, the daylight hours at which hyenas
388 are most active. We deployed the MAB anytime a suitable group of hyenas was located within the territories of our
389 study clans. A suitable group was defined as one containing five or fewer hyenas within 100 m or within visible
390 range that were either walking or resting but not engaged in hunting, border patrol, mating, courtship, or nursing
391 behaviors. We used our research vehicle as a mobile blind to shield the researchers from the view of hyenas while
392 we baited and deployed the MAB on the opposite side of the vehicle from hyenas. We baited the MAB with
393 approximately 200 g of either goat or beef muscle, skin, or organ meat. During familiarization trials we also used
394 full cream milk powder in addition to, or in place of, meat. We deployed the MAB approximately 20 m away from
395 the nearest hyena and after MAB deployment we drove the research vehicle to a distance of 20 to 50 m away from

396 the MAB. We began videotaping immediately after we deployed the MAB and we ended videotaping when we
397 collected the MAB.

398 During familiarization trials we deployed the MAB with the top removed to acclimate subjects to the
399 MAB and allow them to learn to associate the MAB with bait. Familiarization trials began when a hyena came
400 within 5 m of the MAB and ended upon successful food retrieval (a “feed” trial) or when the hyena moved more
401 than 5m away from the MAB for more than 5 minutes. We recored hyenas that approached the MAB, but did not
402 make contact, as not participating in the trial. Average duration of familiarization trials was 11.7 ± 12.3 minutes.

403 If a hyena had a “feed” familiarization trial or successful test trial, and if it had moved at least 5m away
404 from the MAB, we immediately rebaited the MAB for successive testing. We gave hyenas successive trials as long
405 as the testing conditions remained suitable, as described above, or until researchers ran out of bait. We did not
406 administer successive trials following trials where every hyena that participated was unsuccessful because
407 unsuccessful hyenas were those that had moved beyond 5m from the MAB for more than five minutes without
408 opening the MAB and these hyenas were extremely unlikely to spontaneously re-approach the MAB for another
409 trial. On average, hyenas received 1.53 ± 1.25 trials per testing session and completed testing across 6.31 ± 2.58
410 separate sessions (Supplementary Fig. 1). Most sessions were separated by a median of 1 day (mean \pm SD = 24.18
411 ± 60.30 days, range = 0 – 321 days).

412 We divided test trials into four different phases of testing. During Phase 1, we presented the MAB to
413 hyenas with all four doors accessible. After a hyena had reached completion criterion for phase 1, defined by
414 success with the same door in three out of four consecutive trials, it progressed to Phase 2. During Phase 2, we
415 blocked the door used in Phase 1 by bolting it shut. The same criteria for progression applied to subsequent phases
416 until a hyena reached the criteria for progression with all four doors. We gave hyenas trials until they either reached
417 criterion for all four doors or failed five consecutive trials during any phase of testing. We did not include hyenas
418 that participated in less than five trials, of which none were successful, in our analysis. On average, hyenas
419 participated in 9.64 ± 5.61 trials. Hyenas completed phase 1 in 7.43 ± 2.93 trials (N = 72), phase 2 in 3.67 ± 1.11
420 trials (N = 15), phase 3 in 4.08 ± 1.32 trials (N = 13) and phase 4 in 4.25 ± 1.96 trials (N = 12).

421 We aimed to give every hyena two familiarization trials prior to being given the option to participate in
422 test trials. On average we gave hyenas the opportunity to participate in 1.60 ± 1.54 (mean \pm standard deviation)
423 familiarization trials prior to their first phase one trial, but hyenas only fed from the MAB on an average of $0.94 \pm$
424 1.11 familiarization trials prior to their first phase one trial.

425 When we presented a group of hyenas with the MAB, we configured the MAB for the hyena at the most
426 advanced phase of testing. For example, if one hyena in the group had progressed to Phase 3, but all the others
427 were still on Phase 1, we would configure the MAB for the hyena on Phase 3 and block the doors that hyena had
428 used in Phases 1 and 2. Overall, there were only five trials total in which a hyena solved the MAB in a trial during
429 the ‘wrong’ phase of testing by joining a trial where we configured the MAB for a group mate rather than itself.
430 The average ‘trial group size’ per hyena per trial was 3.89 ± 3.71 hyenas (median = 3, range = 1 – 29). We
431 calculated trial group size as a count of all hyenas that participated in a trial by contacting the MAB at any point in
432 time during the trial. Overall, trial group size had a positive and significant effect on hyena participation; hyenas
433 were slightly more likely to contact the box if there were other hyenas contacting the box (Binomial GLMM: $z =$
434 9.19 , $P < 0.001$; Supplementary Table 1). We also examined the effect of ‘overall group size’ which we calculated
435 as a count of all hyenas present within 20 m of the MAB. Overall group size had slightly negative effect on
436 participation (Binomial GLMM: $z = -9.81$, $P < 0.001$; Supplementary Table 1); hyenas were slightly less likely to
437 contact the box if there were more hyenas present within 20 m of the MAB.

438 Overall, we deployed the MAB 483 independent times including both familiarization trials and test trials
439 to 280 different hyenas for a total of 2891 observations. The dataset used in the present analysis only includes test
440 trials from subjects that completed testing by reaching criterion for failure or subjects who had solved the MAB at
441 least once (N = 72 hyenas, n = 694 observations). Of these 72 hyenas, 23 opened the MAB at least once (n = 318
442 trials). On average, we deployed the MAB 120.75 ± 25.80 times to each of our four study clans. In order to
443 identify the 23 solvers, we deployed the MAB an average of 88.5 ± 34.72 (N=4) times in each of our four study
444 clans. In other words, by the 90th deployment on average, we had no new subjects solve the MAB that had not
445 already solved it at least once.

446 **Statistical Analysis**

447 All statistical analyses were performed using the statistical software R [47]. [47]. Here, R values were calculated for
448 subject ID in generalized linear mixed models (GLMMs). The rptR package also allowed us to estimate uncertainty
449 around each point estimate for R via parametric bootstrapping (n = 1000), in which we estimated a standard error, a
450 95% confidence interval, and a P value for each estimate of R. P values were generated using likelihood ratio tests

451 where model fit was compared to a null model with no grouping factor. Here, we report both adjusted R values,
452 calculated as a ratio of the variance explained by subject ID over the residual variance, and conditional R values,
453 calculated as the ratio of the variance explained by subject ID over the total variance, including fixed effects.

454 Before calculating R values for innovative problem-solving ability across doors, we created a global
455 GLMM that included door, age class, sex, and rank as fixed effects and subject ID as a random effect in order to
456 identify factors that might influence innovativeness. We used the glmmTMB package to create all global models
457 [48]. To identify fixed effects of importance, we used the ‘dredge’ function in the R package MuMIn [49]. We built
458 our final model including only the factors with large or significant effects on innovative problem-solving as fixed
459 effects. Dredge identified nine top models for our global model on problem-solving success across doors (Δ AICc <
460 4). No factors were included in all nine top models, but age class was included in the most top models (N=5) and
461 had a large effect that trended towards significance (Supplementary Table 2). Therefore, we included only age class
462 as a fixed effect in our final model.

463 Likewise, before calculating R values for innovative problem-solving ability across trials, we created a
464 global model that included age class, sex, rank, trial number, and phase of testing as fixed effects, and subject ID as
465 a random effect. To identify fixed effects of importance we used the ‘dredge’ function in the R package MuMIn
466 [49]. We fixed trial number and phase of testing for inclusion in all models because we wanted to control for the
467 effects of experience and task difficulty. Dredge identified eight top models with a delta AICc of less than four
468 (Supplementary Table 3). Here, both trial number and phase had significant effects. Once again, age class was only
469 marginally significant but also had the largest effect size (Supplementary Table 3). Therefore we included test trial,
470 phase of testing, and age class in all subsequent models examining problem-solving success across trials (Models
471 2-5.9). In Model 2, phase of testing had a significant negative effect on the likelihood of solving the MAB
472 (GLMM: $\beta = -0.62$, SE = 0.31, $z = -2.01$, P = 0.04) which suggests that later phases of testing, where solutions that
473 hyenas used previously were blocked, were indeed more demanding for hyenas. After controlling for the effect of
474 phase, overall test trial number had a significant positive effect on the likelihood of solving the MAB (GLMM: $\beta =$
475 0.11, SE = 0.06, $z = 1.96$, P = 0.05), indicating that hyenas were more likely to solve the MAB in later than earlier
476 trials (Fig. 6A). The positive effect of trial number could indicate that hyenas were learning how to improve their
477 performance across trials, but this effect might also be biased by some subjects reaching the criterion to end testing
478 (five unsuccessful trials in a row) and dropping out of the subject pool. To test this possibility we created another
479 model where we restricted our dataset to the first ten test trials only for hyenas that had at least 10 trials, and found
480 that test trial still had a significant positive effect on the likelihood of solving the MAB (GLMM: $\beta = 0.30$, SE =
481 0.10, $z = 2.91$, P = 0.004, n = 139 trials, N = 14 hyenas; Fig. 6B).

482 Before calculating R values all models were checked for collinearity by examining variance inflation
483 factors (VIF). Test trial number and phase of testing consistently had VIFs > 4 in most of our models, however, we
484 chose to include both because the high collinearity here is a result of our test protocol; hyenas only progressed to
485 phase four of testing after completing a relatively large number of trials. The main concern with high VIFs is that
486 the estimates error and P values for the collinear factors will be increased; however, both test trial and phase of
487 testing had consistent significant effects, which suggests that this was not a problem in our models. Next, we also
488 examined QQ plot residuals and a histogram of the residuals using the R package DHARMA to confirm that model
489 assumptions about the normality of residuals were not violated.

490 Ethics Statement

491 This work was conducted under research permit no. NACOSTI/P/16/35513/10422, issued by
492 the Kenyan National Commission on Science, Technology and Innovation. The data collection
493 procedure followed here was also approved by the Michigan State University Institutional
494 Animal Care and Use Committee (IACUC): AUF #04/16-050-00. All research procedures were
495 designed to adhere to the American Society of Mammalogists (ASM) Guidelines for the use of
496 wild mammals in research and education [50][50] and to the Association for the Study of
497 Animal Behaviour (ASAB) Ethics Committee and the Animal Behaviour Society (ABS) Animal
498 Care Committee Guidelines for the treatment of animals in behavioural research and teaching
499 [51][51].

500 References

501 1. MacLean, E. L. *et al.* The evolution of self-control. *Proc. Natl. Acad. Sci. U. S. A.* **111**,

- 502 E2140-8 (2014).
- 503 2. Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M. & Holekamp, K. E. Brain
504 size predicts problem-solving ability in mammalian carnivores. *Proc. Natl. Acad. Sci.*
505 **113**, 2532–2537 (2016).
- 506 3. Fristoe, T. S., Iwaniuk, A. N. & Botero, C. A. Big brains stabilize populations and
507 facilitate colonization of variable habitats in birds. *Nat. Ecol. Evol.* **1**, 1706–1715 (2017).
- 508 4. Dunbar, R. I. M. & Shultz, S. Evolution in the Social Brain. *Science (80-.)*. **317**, 1344–
509 1347 (2007).
- 510 5. DeCasien, A. R., Williams, S. A. & Higham, J. P. Primate brain size is predicted by diet
511 but not sociality. *Nat. Ecol. Evol.* **1**, 0112 (2017).
- 512 6. Ashton, B. J., Thornton, A. & Ridley, A. R. An intraspecific appraisal of the social
513 intelligence hypothesis. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170288 (2018).
- 514 7. Cauchoix, M., Hermer, E., Chaine, A. S. & Morand-Ferron, J. Cognition in the field:
515 comparison of reversal learning performance in captive and wild passerines. *Sci. Rep.* **7**,
516 12945 (2017).
- 517 8. Thornton, A., Isden, J. & Madden, J. R. Toward wild psychometrics: linking individual
518 cognitive differences to fitness. *Behav. Ecol.* **25**, 1299–1301 (2014).
- 519 9. Pritchard, D. J., Hurly, T. A., Tello-Ramos, M. C. & Healy, S. D. Why study cognition in
520 the wild (and how to test it)? *J. Exp. Anal. Behav.* **105**, 41–55 (2016).
- 521 10. Boogert, N. J., Madden, J. R., Morand-Ferron, J. & Thornton, A. Measuring and
522 understanding individual differences in cognition. *Philos. Trans. R. Soc. B Biol. Sci.* **373**,
523 20170280 (2018).
- 524 11. Cauchoix, M. *et al.* The repeatability of cognitive performance: a meta-analysis. *Philos.*
525 *Trans. R. Soc. B Biol. Sci.* **373**, 20170281 (2018).
- 526 12. Bell, A. M., Hankison, S. J. & Laskowski, K. L. The repeatability of behaviour: a meta-
527 analysis. *Anim. Behav.* **77**, 771–783 (2009).
- 528 13. Morand-Ferron, J., Cole, E. F. & Quinn, J. L. Studying the evolutionary ecology of
529 cognition in the wild: a review of practical and conceptual challenges. *Biol. Rev. Camb.*
530 *Philos. Soc.* (2015) doi:10.1111/brv.12174.
- 531 14. van Horik, J. O. & Madden, J. R. A problem with problem solving: motivational traits,
532 but not cognition, predict success on novel operant foraging tasks. *Anim. Behav.* **114**,
533 189–198 (2016).
- 534 15. *Animal Innovation*. (Oxford University Press, 2003).
535 doi:10.1093/acprof:oso/9780198526223.001.0001.
- 536 16. Shaw, R. C. Testing cognition in the wild: factors affecting performance and individual
537 consistency in two measures of avian cognition. *Behav. Processes* **134**, 31–36 (2017).
- 538 17. Cole, E. F., Cram, D. L. & Quinn, J. L. Individual variation in spontaneous problem-
539 solving performance among wild great tits. *Anim. Behav.* **81**, 491–498 (2011).
- 540 18. McCune, K. B., Jablonski, P., Lee, S. & Ha, R. R. Captive jays exhibit reduced problem-
541 solving performance compared to wild conspecifics. *R. Soc. Open Sci.* **6**, 181311 (2019).
- 542 19. Ashton, B. J., Ridley, A. R., Edwards, E. K. & Thornton, A. Cognitive performance is
543 linked to group size and affects fitness in australian magpies. *Nature* **554**, 364–367
544 (2018).
- 545 20. Holekamp, K. E., Sakai, S. T. & Lundrigan, B. L. The spotted hyena (*Crocuta crocuta*) as
546 a model system for study of the evolution of intelligence. *J. Mammal.* **88**, 545–554
547 (2007).

- 548 21. Benson-Amram, S. & Holekamp, K. E. Innovative problem solving by wild spotted
549 hyenas. *Proc. R. Soc. B Biol. Sci.* **279**, 4087–95 (2012).
- 550 22. Benson-Amram, S., Weldele, M. L. & Holekamp, K. E. A comparison of innovative
551 problem-solving abilities between wild and captive spotted hyaenas, *Crocuta crocuta*.
552 *Anim. Behav.* **85**, 349–356 (2013).
- 553 23. Johnson-Ulrich, L., Johnson-Ulrich, Z. & Holekamp, K. Proactive behavior, but not
554 inhibitory control, predicts repeated innovation by spotted hyenas tested with a multi-
555 access box. *Anim. Cogn.* **21**, 379–392 (2018).
- 556 24. Johnson-Ulrich, L., Benson-Amram, S. & Holekamp, K. E. Fitness Consequences of
557 Innovation in Spotted Hyenas. *Front. Ecol. Evol.* **7**, 443 (2019).
- 558 25. Nakagawa, S. & Schielzeth, H. Repeatability for Gaussian and non-Gaussian data: A
559 practical guide for biologists. *Biol. Rev.* **85**, 935–956 (2010).
- 560 26. Griffin, A. S., Guillette, L. M. & Healy, S. D. Cognition and personality: an analysis of
561 an emerging field. *Trends Ecol. Evol.* **30**, 207–14 (2015).
- 562 27. Stoffel, M. A., Nakagawa, S. & Schielzeth, H. rptR: repeatability estimation and variance
563 decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **8**, 1639–
564 1644 (2017).
- 565 28. Huebner, F., Fichtel, C. & Kappeler, P. M. Linking cognition with fitness in a wild
566 primate: fitness correlates of problem-solving performance and spatial learning ability.
567 *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170295 (2018).
- 568 29. Thornton, A. & Samson, J. Innovative problem solving in wild meerkats. *Anim. Behav.*
569 **83**, 1459–1468 (2012).
- 570 30. Manrique, H. M., Völter, C. J. & Call, J. Repeated innovation in great apes. *Anim. Behav.*
571 **85**, 195–202 (2013).
- 572 31. Huebner, F. & Fichtel, C. Innovation and behavioral flexibility in wild redfronted lemurs
573 (*Eulemur rufifrons*). *Anim. Cogn.* **18**, 777–87 (2015).
- 574 32. Borrego, N. & Dowling, B. Lions (*Panthera leo*) solve, learn, and remember a novel
575 resource acquisition problem. *Anim. Cogn.* **19**, 1019–1025 (2016).
- 576 33. Chow, P. K. Y., Lea, S. E. G. & Leaver, L. A. How practice makes perfect: the role of
577 persistence, flexibility and learning in problem-solving efficiency. *Anim. Behav.* **112**,
578 273–283 (2016).
- 579 34. Lermite, F., Peneaux, C. & Griffin, A. S. Personality and problem-solving in common
580 mynas (*Acridotheres tristis*). *Behav. Processes* **134**, 87–94 (2017).
- 581 35. Ducatez, S., Audet, J. N. & Lefebvre, L. Problem-solving and learning in Carib grackles:
582 individuals show a consistent speed–accuracy trade-off. *Anim. Cogn.* **18**, 485–496
583 (2015).
- 584 36. Griffin, A. S. & Guez, D. Innovation and problem solving: A review of common
585 mechanisms. *Behav. Processes* **109**, 121–134 (2014).
- 586 37. Diquelou, M. C., Griffin, A. S. & Sol, D. The role of motor diversity in foraging
587 innovations: a cross-species comparison in urban birds. *Behav. Ecol.* **27**, 584–591 (2015).
- 588 38. Chow, P. K. Y., Lea, S. E. G., Hempel de Ibarra, N. & Robert, T. How to stay perfect: the
589 role of memory and behavioural traits in an experienced problem and a similar problem.
590 *Anim. Cogn.* **20**, 941–952 (2017).
- 591 39. Smith, J. E. & Holekamp, K. E. Spotted Hyenas. *Encycl. Anim. Behav.* **3**, 190–208
592 (2018).
- 593 40. Rowe, C. & Healy, S. D. Measuring variation in cognition. *Behav. Ecol.* **25**, 1287–1292

- 594 (2014).
- 595 41. Reader, S. M., Morand-Ferron, J. & Flynn, E. Animal and human innovation: novel
596 problems and novel solutions. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150182 (2016).
- 597 42. Van Horn, R. C., McElhinny, T. L. & Holekamp, K. E. Age Estimation and Dispersal in
598 the Spotted Hyena (*Crocuta crocuta*). *J. Mammal.* **84**, 1019–1030 (2003).
- 599 43. Engh, A. L., Esch, K., Smale, L. & Holekamp, K. Mechanisms of maternal rank
600 ‘inheritance’ in the spotted hyaena, *Crocuta crocuta*. *Anim. Behav.* **60**, 323–332 (2000).
- 601 44. Holekamp, K. E. & Dloniak, S. M. Intraspecific Variation in the Behavioral Ecology of a
602 Tropical Carnivore, the Spotted Hyena. in *Behavioral ecology of tropical animals* vol. 42
603 189–229 (Elsevier, 2010).
- 604 45. Holekamp, K. E., Smale, L. & Szykman, M. Rank and reproduction in the female spotted
605 hyaena. *Reproduction* **108**, 229–237 (1996).
- 606 46. Auersperg, A. M. I., Gajdon, G. K. & von Bayern, A. M. P. A new approach to
607 comparing problem solving, flexibility and innovation. *Commun. Integr. Biol.* **5**, 140–145
608 (2012).
- 609 47. R Core Team. R: A language and environment for statistical computing. (2019).
- 610 48. Brooks, Mollie, E. *et al.* glmmTMB Balances Speed and Flexibility Among Packages for
611 Zero-inflated Generalized Linear Mixed Modeling. *R J.* **9**, 378 (2017).
- 612 49. Bartoń, K. MuMIn: Multi-Model Inference. (2018).
- 613 50. Sikes, R. S. 2016 Guidelines of the American Society of Mammalogists for the use of
614 wild mammals in research and education. *J. Mammal.* **97**, 663–688 (2016).
- 615 51. Behaviour, A. Guidelines for the treatment of animals in behavioural research and
616 teaching. *Anim. Behav.* **123**, I–IX (2017).
- 617

618 **Acknowledgements**

619 We thank the Kenyan National Commission for Science, Technology and Innovation, the Kenya
620 Wildlife Service, the Narok County Government, the Mara Conservancy, Brian Heath, James
621 Sindiyo, and the Naibosho Conservancy for permissions to conduct this research. This work was
622 supported by NSF grants OISE1853934 and IOS 1755089 and by an NSF graduate research
623 fellowship to LJU.

624 **Author contributions statement**

625 LJU, DZH & KEH conceived the experiment, LJU conducted the experiments and analyzed the
626 data with support from KEH. All authors wrote and reviewed the manuscript.

627 **Competing interests statement**

628 The authors declare that they have no competing financial and/or non-financial interests in
629 relation to the work described here.

630 **Data availability statement**

631 All data and R code used in this manuscript are available in the KNB Data Repository [ID
632 inserted here].