INTRASPECIFIC VARIATION IN BEHAVIORAL FLEXIBILITY IN SPOTTED HYENAS

By

Lily Johnson-Ulrich

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Integrative Biology – Doctor of Philosophy
Ecology, Evolutionary Biology, and Behavior – Dual Major

2020

ABSTRACT

INTRASPECIFIC VARIATION IN BEHAVIORAL FLEXIBILITY IN SPOTTED HYENAS

By

Lily Johnson-Ulrich

The question of how intelligence evolves across the animal kingdom remains unresolved despite more than a century of intensive research. The Social Intelligence Hypothesis (SIH) predicts a primarily social function of intellect whereas the Cognitive Buffer Hypothesis (CBH) suggests that intelligence is most adaptive for dealing with environmental novelty and change. Recently, there has been growing recognition of the value of intraspecific studies of wild, free-ranging populations because they allow researchers to examine the correlates of cognitive variation in an ecologically valid context, and also to examine the fitness consequences of this variation. In my dissertation, I examine the causes and consequences of intraspecific variation in behavioral flexibility in wild, free-ranging spotted hyena (*Crocuta crocuta*) populations. Behavioral flexibility is the ability to adaptively modify a behavior as current circumstances demand and, akin to fluid intelligence in humans, behavioral flexibility is thought to be an important aspect of intelligence across non-human animals. Spotted hyenas show a great deal of intraspecific variation in behavioral flexibility and persist under a diversity of socio-ecological conditions, making them an ideal model organism for testing hypotheses about the evolution of intelligence with the intraspecific approach. In Chapter One, I review hypotheses for the evolution of intelligence and the methods used to measure intelligence across animal taxa. Next, in Chapters Two and Three, I present evidence for the validity and reliability of apparatus used to measure behavioral flexibility by examining the individual traits that predict high behavioral flexibility and individual consistency in performance. Next, in Chapter Four, I examine the social correlates of behavioral flexibility. My results reveal that socially low-ranking hyenas that live in larger groups have greater behavioral flexibility and that hyenas of any rank that grow up in larger cohorts have greater behavioral flexibility, a result that supports predictions of the SIH. In Chapter Five, I examine the ecological correlates of behavioral flexibility by comparing behavioral flexibility across populations of hyenas experiencing varying degrees of urbanization. Urbanization causes widespread environmental change and introduces a high degree of novelty into habitats; however, my results reveal that the degree of urbanization is negatively correlated with behavioral flexibility, which challenges the CBH. Finally, I examine the fitness consequences of behavioral flexibility in Chapter Six. My results reveal that behaviorally flexible females have lower offspring survivorship but produce significantly more cubs annually than behaviorally inflexible females, which suggests there may be important trade-offs with regards to behavioral flexibility.

In sum, my research demonstrates that the intraspecific approach is a powerful tool for understanding how and why behavioral flexibility evolves. My results generally support the SIH and suggest that social living is a key correlate of behavioral flexibility in spotted hyenas. However, it appears that high behavioral flexibility in spotted hyenas may have also been an important factor in facilitating their adaptation to urban environments. Urbanization is putting an increasing number of species at risk globally, and understanding how or why some animals adapt, while others go extinct, is crucial for wildlife conservation.

ACKNOWLEDGEMENTS

I would like to thank the Kenyan National Commission for Science, Technology and Innovation, the Kenya Wildlife Service, the Narok County Government, the Mara Conservancy, James Sindiyo, and the Naibosho Conservancy for permissions to conduct this research in Kenya. Special thanks to Brian Heath, manager of the Mara Conservancy, for his support of my research and continuous support of the Mara Hyena Project. I won't forget the game drive you took myself and Emily Thomas on where we spotted an albino spotted hyena cub that was less than three months old. I would also like to thank Dr. Gidey Yirga and Mekelle University for their support and assistance in conducting this research in Ethiopia. Dr. Gidey, an expert on urban hyenas, was an enthusiastic collaborator and gracious host in Mekelle, Ethiopia. This work was supported by NSF grants OISE1556407, OISE1853934, DEB1353110 and IOS1755089 to Dr. Kay E. Holekamp and by an NSF graduate research fellowship to me. I was also supported by grants from the International Society for Behavioral Ecology, the Animal Behaviour Society, and the Society for Integrative and Comparative Biology. Furthermore, I received fellowships from the Council of Graduate Students, College of Natural Sciences, the Integrative Biology Department, and the program in Ecology, Evolutionary Biology, and Behavior at Michigan State University. This work was also supported in part by the BEACON Center for the Study of Evolution in Action, funded by National Science Foundation Grant OIA0939454.

This dissertation would not have been possible without the support, guidance, and direction of my advisor, Dr. Kay E. Holekamp. Kay did a fantastic job of gently nudging me in the right direction when I was uncertain of which steps to take and giving her honest, straight to the point, opinion when I needed advice. Kay, I will always be grateful for your invaluable guidance, for the example you set on how to conduct research by combining expansive scientific expertise with intimate knowledge of spotted hyena ecology, and for giving me the opportunity to study spotted hyenas myself. Likewise, my committee members Dr. Laura Smale, Dr. Heather Eisthen, Dr. Zach Hambrick, and previously Dr. Ashlee Rowe, provided enormously helpful advice over the years that has strengthened my scientific thinking and writing. I also thank Andrew Dennhardt at the MSU's Center for Statistical Training and Consulting (CSTAT) for his considerable help with my statistics and model selection. I thank Roy Bailiff at MSU Engineering for helping build the cylinder and multi-access box apparatus. I would also like the thank the numerous support staff in the Department of Integrative Biology, EEBB, and BEACON at MSU for their quick and helpful responses to all of my queries.

The MSU Hyena Lab and the Mara Hyena Project has been a close community of support and I'm thankful for the many friendships, collaborations, and co-authorships that were born out of the Hyena Lab including graduate students and postdocs: Dr. David Green, Dr. Julia Greenberg, Dr. Julie Turner, Dr. Agathe Laurence, Dr. Sarah Jones, Tracy Montgomery, Kenna Lehmann, Dr. Eli Strauss, Dr. Nora Lewin, Kevin McCormick, Connie Rojas, Maggie Sawdy, Olivia Spagnuolo, and Julie Jarvey, and research assistants: Wes Binder, Emily Thomas, Phoebe Parker-Shames, Hadley Courad (who also married my wife and I), Amy Fontaine, Rebecca LaFleur, Morgan Lucot, Mary Kearney, Leah McTigue, Jack Grady, Mike Kowalski, Emily Nonnamaker, and Kecil Johns. A special thanks to research assistant Robyn Strong who not only assisted me in Kenya and coded hundreds of hours of videos, but also traveled with me to Ethiopia for 3 months where we lived in very remote city of Mekelle and struggled together to adjust to the culture and food while spending our nights collecting data on nocturnal hyenas with infrared cameras. Undergraduates Paige Barnes, Dan Macina, Sarah Vanarsdale-Hoot, McKenna Barlass, and Kaelyn Seestadt all spent countless hours coding hyena videos and extracting data.

The project's staff in Kenya also provided amazing support and friendship during my two and a half years living in Kenya. In Serena camp I thank Philimon Naiguran, Moses Naiguran, and George Kilenyet. In Talek camp I thank Joseph Kamaamia, Jackson Kamaamia, Dickson Pion, Sasine Rotiken, Samuel Kamaamia, Steven Karkar, and Lesinko Naurori. Philimon, thanks for being a great Kiswahili teacher and gracious host. A special thanks to Kenyan research assistant Benson Pion, your experience and skills were of crucial importance for my own data collection and life in Kenya. I can't thank you enough for the number of times you've pulled me out the mud, for your gentle teaching method as you helped me learn Talek hyenas, your astute observations as you located and identified hyenas for me to test with the cylinder and the box, and your friendship. I fondly remember taking breaks from data collection to identify birds and I owe you for finding that Sulphur-breasted bush shrike (your note is still in my bird book)!
I would also like to thank the Maji researchers, Amanda Subaluski and Chris Dutton, with whom we shared our Serena campsite with. Chris and Amanda, thanks for being a warm and welcoming presence in camp with your offer of friendship, pizza, and sundowners to all of the Hyena RAs and graduate students that passed through camp. The staff at Serena Lodge are also owed a thank you for letting us use their guest facilities, borrow their Wi-Fi and power, shop for supplies at their staff canteen, and letting us share fun hyena facts to their guests once a week.

Last, I'd like to thank my family, my pets, and my wife for their emotional support during the undertaking of my dissertation. I thank my father, Russdee, for inspiring and facilitating a love of science in myself. I thank my mother Katrina, for inspiring and facilitating a love of travel and culture in myself. Special thanks to my twin sister, Dr. Zoe Johnson-Ulrich, who has probably read every single rough draft of everything important that I've ever written. My pet cat, Bindi, started her journey with me when I was an undergraduate in Coral Gables, Florida and eventually traveled with me to Lansing, Michigan where she was my study buddy during my first two years of grad school and where she taught me how to clicker train her for high fives and target boops. Askari became a loving companion in the Talek camp site of Kenya where she immediately staked a claim in everyone's hearts. A little brown Maasai dog; I've never met a happier, free-loving, and playful spirit. Bindi and Askari, I wish you could be here with me today. Thank you to my wife Bethany, who I met the summer I started at MSU. As a fifth year graduate student in Chemistry, your first 'how to survive grad school' advice to me was to suggest me that if I was going to get hit by a car while I biked to campus, I'd better try to get hit by one of those Ferrari's. Bethany helped me navigate the uncertainties of graduate study and has been a constant source of support. With Bethany I also became the step-owner to two new pets, Peavey and Nim. Peavey was a fantastic lap cat during my last years of graduate school and Nim has been a source of joy and adventure as I've watched her grow from a young six-month-old puppy to a quiet and affectionate adult dog. Nim has always made sure that I absolutely, never, ever, get left behind.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

GENERAL INTRODUCTION

Hypotheses for the evolution of intelligence

Despite wide interest, the evolution of intelligence across animals continues to confound researchers today (Dunbar & Shultz, 2017; Rosati, 2017). Originally driven by an interest in understanding how human intelligence evolved, researchers are interested in determining whether there any general trends in the evolution of intelligence across all animals. Several hypotheses have been put forth to explain the evolution of general intelligence and large brains in animals, most of which can be divided into either social or ecological hypotheses. Social hypotheses focus on the role of group living (Dunbar, 1998), forming social relationships (Dunbar & Shultz, 2007; Kudo & Dunbar, 2001), cultural transmission of knowledge (Humphrey, 1976; van Schaik & Burkart, 2011), and social cognition in animal lives (Humphrey, 1976; Jolly, 1966). Among the most popular of social hypotheses is the Social Intelligence Hypothesis (SIH), which suggests that social intelligence evolved as a direct result of the cognitive demands imposed by living in complex social groups, and that general intelligence results as by-product of selection on social intelligence (Ashton, Thornton, & Ridley, 2018; Byrne, 1995b; Dunbar, 1998; Dunbar & Shultz, 2007).

Ecological hypotheses, on the other hand, suggest that factors such as extractive foraging (King, 1986; Parker, 2015), tool-use or causal reasoning (Byrne, 1997), environmental change (Sol, Duncan, Blackburn, Cassey, & Lefebvre, 2005), habitat complexity and spatial-mapping (Milton, 1981), or dietary breadth (Clutton-Brock & Harvey, 1980) may have selected for intelligence. Like the social hypotheses, these hypotheses attempt to pinpoint a particular aspect of animal lives' that is the most cognitively demanding. However, because of this specificity, these hypotheses largely fail to explain the presence of domain-general abilities such as general intelligence. One popular hypothesis that attempts to explain the presence of general intelligence is the Cognitive Buffer Hypothesis (CBH). The CBH suggests that general intelligence, by increasing behavioral flexibility, buffers individuals against the negative effects of occasional catastrophic events (Allman, McLaughlin, & Hakeem, 1993). The CBH also links larger brains to longer lifespans, because longer lived animals may be more likely to encounter such catastrophic events. A long-lifespan may also be linked to larger brains for the functional reason that brain tissue is energetically expensive and larger brains require a great deal more time to develop than smaller brains. Consequently, only species that can afford the high price of large brains will evolve large brains (Isler & van Schaik, 2009; Liao et al., 2016; Weisbecker, Blomberg, Goldizen, Brown, & Fisher, 2015).

Overall, both the SIH and CBH have garnered significant support, but important questions still remain. In primates both social complexity and ecological complexity appear highly correlated, making it difficult to disentangle their respective contributions to the evolution of intelligence and large brains in primates (Reader et al., 2011; Rosati, 2017). Outside of primates some studies have found positive correlations between social complexity and brain size, but an important weaknesses of the SIH is its inability to adequately define social complexity and its inability to explain grade shifts in brain size among taxa (van Schaik, Isler, & Burkart, 2012). For example, within Hyaenidae spotted hyenas (*Crocuta crocuta*; hereafter 'hyenas') live in the largest groups and have the largest brains which supports predictions of the SIH, but hyenas are relatively small-brained compared to primates despite sharing very similar social structures and sharing many of the same social cognitive abilities (Holekamp, 2007). These grade shifts suggest that challenges outside the social realm may also be important and may partially result from the fact that brain size evolution is severely limited by things such as body size and energetic constraints. Likewise, the CBH is strongly supported by work showing that animals with larger brains are more likely to invade novel and rapidly changing environments (Amiel, Tingley, & Shine, 2011; Sol, Bacher, Reader, & Lefebvre, 2008; Sol, Timmermans, & Lefebvre, 2002), however, it appears that many of the species used in these studies evolved their large brains prior to invading novel environments and that brain size is just as likely to evolve in climatically stable environments (Fristoe, Iwaniuk, & Botero, 2017). As a consequence of the conflicting results produced by these phylogenetic studies, researchers have recently begun using the intraspecific approach (Ashton, Thornton, et al., 2018). The intraspecific approach allows researchers to identify aspects of extant species' environment that are cognitively demanding by correlating variation in cognitive abilities with variation in socio-ecological demands.

Defining and measuring intelligence

One of the largest challenges with studying the evolution of intelligence is defining and measuring intelligence across diverse taxa. Animals exhibit a diverse array of cognitive abilities and it is likely that all animals are equipped with the cognitive skills they require to survive in their niche, i.e. domain-specific intelligence (Kamil, 1987; Shettleworth, 1998). However, a great deal of research has also uncovered evidence for domain-general intelligence in animals (Burkart, Schubiger, & van Schaik, 2017). General intelligence (g) is broadly defined in both humans and non-human animals as the ability to acquire new information, integrate information from multiple sources, and use this information in flexible, novel, and adaptive ways (Burkart et al., 2017). It is often g, rather than specialized cognitive abilities, that researchers are concerned with when discussing the evolution of intelligence. In humans, g is a fairly robust and well-studied phenomenon that is typically measured by estimating the strength of a single latent factor underlying cognitive performance across multiple domains (Gottfredson, 1997). In animals, g is measured using several different approaches, outlined below, which each have their own weaknesses. First, a significant portion of the literature on the evolution of intelligence in animals uses various measures of brain size as a proxy for intelligence (Barton & Dunbar, 1997; Fristoe & Botero, 2019; Harvey & Krebs, 1990; Powell, Isler, & Barton, 2017). The volume of brain gray matter is robustly correlated with g in humans (Nave, Jung, Karlsson Linnér, Kable, & Koellinger, 2019) and an increasing amount of literature has linked performance on various cognitive tasks to different measures of brain size in non-human animals (Benson-Amram, Dantzer, Stricker, Swanson, & Holekamp, 2016; Herculano-Houzel, 2017; Herczeg et al., 2019; Horschler et al., 2019; Kotrschal et al., 2013b; Lindsay, Houck, Giuliano, & Day, 2015; Shultz & Dunbar, 2010). However, this approach suffers from several well-examined weaknesses in the study of non-human animals (Broadway, Samuelson, Christopher, Jett, & Lyn, 2017; Deaner, Nunn, & van Schaik, 2000; S. D Healy & Rowe, 2007; Susan D. Healy & Rowe, 2013; Roth & Dicke, 2005). For example, brain size is strongly positively correlated with body size across most animals and, while most researchers do attempt to control for this allometry, they do so using a wide diversity of techniques and the specific techniques used influence the evolutionary conclusions drawn (Deaner et al., 2000). In addition, brains differ not just in size but also in the size, shape, and composition of different parts of the brain (i.e. mosaic brain evolution (Smaers & Soligo, 2013)). Furthermore, even similarly sized brains may contain vastly different numbers of, or differently organized, neurons and other brain cells (Herculano-Houzel, 2011), which calls into question the validity of comparing brain size across diverse taxa (Willemet, 2013). Ultimately, there is still a great deal of debate about the ‘best’ way to measure variation in brain size (Broadway et al., 2017; Deaner, Isler, Burkart, & van Schaik, 2007; Herculano-Houzel, 2017). Unfortunately, all hypotheses for the evolution of intelligence in animals are founded on studies that used brain size as a proxy for intelligence (reviewed in Mitchell, 2016). Some researchers have attempted to measure g directly in animals by giving subjects batteries of cognitive tests and calculating a common factor underlying results on these tests (Burkart et al., 2017; Damerius et al., 2019; Shaw & Schmelz, 2017; Woodley of Menie, Fernandes, & Peñaherrera Aguirre, 2017). This approach closely mirrors the approach used to measure g in humans and holds great promise for measuring g in animals (but see Shuker, Barrett, Dickins, Scott-Phillips, & Barton, 2017; Van Horik, Langley, Whiteside, Laker, & Madden, 2018). However, giving batteries of tests to animal subjects, especially in the wild, is difficult and time consuming (Shaw, 2017). Instead, many researchers choose to measure a single cognitive trait that is linked to or correlated with g. In humans, general intelligence is described as being made up of two different types of intelligence: fluid intelligence and crystallized intelligence. While crystallized intelligence encompasses knowledge that individuals have accumulated over time, fluid intelligence is an individual's ability to think logically and learn new things (Burkart et al., 2017). The concept of fluid intelligence is nearly synonymous with the concept of behavioral flexibility in the field of animal behavior. Behavioral flexibility is the ability to adaptively modify a behavior as current circumstances demand (Jones, 2005; Lea, Chow, Leaver, & McLaren, 2020). Cognitive abilities that require a high degree of behavioral flexibility also tend to strongly correlate with g (Diamond, 2013; Reader et al., 2011; Roth & Dicke, 2005).

GENERAL METHODS

In this dissertation, I measure two cognitive traits that require high behavioral flexibility, innovation and inhibitory control. Innovation is the ability to solve a novel problem or use a novel behavior to solve a familiar problem (Reader & Laland, 2003). Innovation is a key aspect of culture, and is related to general intelligence in both humans and non-human animals (Reader, Morand-Ferron, & Flynn, 2016). Inhibitory control is an executive function also related to general intelligence in humans and animals (Burkart et al., 2017; Diamond, 2013; Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2007). It is the ability to resist performing a prepotent, but unrewarding, behavior in favor of a different behavior (Kabadayi, Bobrowicz, & Osvath, 2018; MacLean et al., 2014). Both innovation and inhibitory control are thought to be good measures of behavioral flexibility because they are both related to the ability to modify behavior when necessary in a flexible manner.

Figure 1.1. The multi-access box (MAB). This is a metal box that is baited with meat. The meat may be obtained by opening one of four doors located on each side of the box. Each of the four doors requires a different motor pattern to open. This figure shows a subadult hyena grasping the door knob in its mouth prior to pulling this door open. Holes are drilled into the walls of the MAB to allow the scent of bait to reach the hyena.

I tested innovation and inhibitory control using a multi-access box (MAB) (Auersperg, Gajdon, & von Bayern, 2012). A MAB is a novel extractive foraging task where subjects may use one of several possible doors to obtain bait from the inside. The MAB used in this dissertation had four different doors (Figure 1.1). Each of the four doors requires a different motor pattern to open and therefore each represents a unique innovation for subjects to solve. Subjects are scored on how many of the four doors they can reliably learn to open. Thus, achieving a high score requires both repeated innovation and repeated use of each innovation. The multi-access box also requires inhibitory control, once a subject has learned one door, that door is fastened shut to force the subject to find a new solution. To open a new door subjects must inhibit using the behaviors that were successful for doors opened earlier.

Figure 1.2. The cylinder task. This is a transparent cylinder that is baited with meat. Subjects may retrieve the meat through openings on either end of the tube. Subjects are tested on their ability to retrieve the meat without bumping into the wall of the cylinder. Holes are drilled into the wall of the cylinder to prevent the scent of bait from leading hyenas to the openings on either end.

I also tested inhibitory control independently of the multi-access box using the 'cylinder task' (Figure 1.2), a detour reaching paradigm (Kabadayi et al., 2018). During testing with the cylinder task subjects are presented with a transparent cylinder that rests horizontally on the ground with bait placed in the center. Subjects must inhibit going straight for the bait and instead detour to openings on either end of the cylinder.

Spotted hyenas have been well-established as an excellent model organism for studying the evolution of intelligence (Holekamp, Sakai, & Lundrigan, 2007). They exhibit convergent social complexity and social intelligence with cercopithicine primates (Holekamp, 2007), compared to other carnivores their diet is relatively flexible (Holekamp & Dloniak, 2010), and they inhabit a diverse array of habitats, including urban environments, across all of sub-Saharan Africa (Holekamp & Dloniak, 2010; Yirga Abay, Bauer, Gebrihiwot, & Deckers, 2010). In short, spotted hyenas experience many of the socio-ecological variables hypothesized to favor the evolution of intelligence.

Although much research on the evolution of intelligence has been conducted using captive animals, measuring cognition in wild populations has several distinct benefits (Susan D Healy & Rowe, 2014; Shaw, 2017; A. Thornton, Isden, & Madden, 2014). By measuring cognitive abilities in wild extant species researchers can examine the various factors that might correlate with the expression or development of such abilities in a socio-ecologically valid context. In addition, field studies allow researchers to measure the fitness consequences of variation in cognitive abilities and determine whether natural selection acts on different cognitive abilities. A long-term field study of spotted hyenas in the Maasai Mara National reserve, Kenya, run by Dr. Kay Holekamp, collected detailed and unique data on the behavior, demography, ancestry, genetics, and hormonal state of hundreds of hyenas. This long-term study allowed me to examine the environmental, social, and individual factors affecting cognition in wild hyenas.

Ethics statement

The data collection procedure followed here was approved by the Michigan State University Institutional Animal Care and Use Committee (IACUC): AUF #04/16-050-00. All research procedures were designed to adhere to the American Society of Mammalogists (ASM) Guidelines for the use of wild mammals in research and education (Sikes, 2016) and to the Association for the Study of Animal Behaviour (ASAB) Ethics Committee and the Animal Behaviour Society (ABS) Animal Care Committee Guidelines for the treatment of animals in behavioral research and teaching (Behaviour, 2017). Utmost care was taken to minimize the impact of our presence on hyenas. While hyenas may have experienced some distress from neophobia during initial trials with apparatus, the bait that was used as a reward usually offset this neophobia quite rapidly and research suggests that working to obtain food from baited apparatus is an enriching experience for captive animals (Clark, 2017; Langbein, Siebert, & Nürnberg, 2009; McGowan, Robbins, Alldredge, & Newberry). All procedures in captive hyenas were reviewed and approved for each zoological institutions' ethical guidelines. Captive hyenas were not fasted prior to testing; instead, testing took place during normal mealtimes for hyenas and apparatus were baited with food that was part of their normal diet. Participation by captive hyenas was voluntary and hyenas that chose not to participate were allowed to leave the testing enclosure after 15 minutes had elapsed. Testing was discontinued for captive hyenas who failed to participate four times in a row. All wild hyenas were part of a population of spotted hyaenas that are extremely well-habituated to the presence of research vehicles and tourist vehicles. Hyena participation was completely voluntary because hyenas could come and go as they pleased. Research vehicles were used as mobile blinds to hide researchers from view of all wild spotted hyaenas during deployment and collection of apparatus. Testing was discontinued after five trials for any subjects that continued to show any neophobia towards apparatus.

OVERVIEW OF DISSERTATION CHAPTERS

In this dissertation, I examined intraspecific variation in behavioral flexibility in spotted hyenas in order to test hypotheses for the evolution of intelligence. I first examined the validity and reliability of methods used to test innovation in hyenas (Chapters Two and Three). Second, I tested hypotheses about the evolution of intelligence in hyenas by examining the socio-ecological correlates of variation in innovation and inhibitory control and their fitness consequences (Chapters Four, Five and Six).

The MAB used to measure innovation was a novel apparatus designed by myself and my sister Zoe Johnson-Ulrich to measure innovation in carnivore species. Because ‘non-cognitive’ factors may influence individual performance on tests designed to measure behavioral flexibility (Rowe & Healy, 2014), in Chapter Two, I first tested innovation in captive hyenas in order to examine the validity of using a MAB for testing innovation. I examined the behavioral traits thought to underlie successful innovation and found evidence that a syndrome of ‘proactive’ traits predicted innovation scores measured with the MAB, but inhibitory control scores measured with the cylinder task did not predict innovation scores. The link between innovation and proactivity suggests that innovation may be closely linked to personality. However, these proactive traits closely mirror the traits that predict spontaneous innovation in wild animals. Ultimately, I concluded that the MAB is a valid measure of innovative abilities.

Next, in order to test evolutionary hypotheses, I measured innovation and inhibitory control in wild hyenas. Testing cognition in the wild has many advantages, but there are also a great number of extraneous or confounding variables present when conducting cognitive tests in the wild (Shaw, 2017). Therefore, in Chapter Three, in order test the assumption that hyenas' scores with the MAB reflect a stable measure of innovativeness I tested the reliability of MAB scores with wild hyenas. Here, I presented evidence that, despite the numerous challenges of testing wild subjects, innovativeness, as measured with the MAB, appears to be a highly reliable trait in hyenas across time and contexts.

Spotted hyenas were previously established as a good model system for testing the SIH and in Chapter Four, I tested predictions of the SIH using the cylinder task with wild hyenas. In this chapter I tested the prediction that the changing social demands created by variation in sex, rank, and group size would affect inhibitory control. I found evidence showing that both rank and group size were positively correlated with inhibitory control in wild hyenas, a result that supports predictions of the SIH.

Next, because ecological demands are also thought to be important in animal's lives I next tested predictions of the CBH using the MAB wild hyenas. In this chapter I compare innovative ability across three different populations of wild hyenas that are experiencing different degrees of environmental novelty and change. Here, I present evidence that hyenas experiencing a high degree of novelty or a high degree of environmental change did not have greater innovative abilities. Instead, hyenas living in the most stable environment, one much like the hyenas' ancestral environment of the Pleistocene, had the highest levels of innovativeness, a result that challenges the hypothesis that novel environments do not themselves favor greater innovation. However, earlier version of the CBH largely focused on the role of innovation in buffering mortality from any cause, not just environmental change.

Therefore, in Chapter Six, I examined the relationship between survival and innovativeness using a single-access puzzle box similar to the MAB described above. This single-access box was presented to wild hyenas to test innovation in earlier studies (Benson-Amram, Heinen, Gessner, Weldele, & Holekamp, 2014; Benson-Amram & Holekamp, 2012). Here, I compared offspring survival for innovative and non-innovative female hyenas. In addition, if innovativeness facilitates survival, it might also facilitate greater reproductive success. I found that innovative female hyenas had lower offspring survival, but higher annual reproductive success compared to non-innovative hyenas which suggests that there may be trade-offs with regards to innovativeness.

In sum, this dissertation investigates the following questions: what is intelligence, how do we test it, and how does it evolve? Here, I attempt to provide part of the answer to these questions by testing the intelligence of wild spotted hyenas and comparing their intelligence to individual, social, and environmental factors. Overall, results from my dissertation suggest that both individual factors and social factors appear to be far more important than environmental variables, such as urbanization, for the expression of behavioral flexibility in spotted hyenas. The following chapters have already been, or will soon be, published as scholarly articles in scientific journals as part of a collaborative effort with other researchers. Therefore, I will use the term ‘we’ in these chapters to acknowledge these collaborations.

CHAPTER TWO

PROACTIVE BEHAVIOR, BUT NOT INHIBITORY CONTROL, PREDICTS REPEATED INNOVATION BY SPOTTED HYENAS TESTED WITH A MULTI-ACCESS BOX

Johnson-Ulrich, L., Johnson-Ulrich, Z., & Holekamp, K. (2018). Proactive behavior, but not inhibitory control, predicts repeated innovation by spotted hyenas tested with a multi-access box. *Animal Cognition*, *21*(3), 379–392. https://doi.org/10.1007/s10071-018-1174-2

ABSTRACT

Innovation is widely linked to cognitive ability, brain size, and adaptation to novel conditions. However, successful innovation appears to be influenced by both cognitive factors, such as inhibitory control, and non-cognitive behavioral traits. We used a multi-access box (MAB) paradigm to measure repeated innovation, the number of unique innovations learned across trials, by 10 captive spotted hyenas (*Crocuta crocuta*). Spotted hyenas are highly innovative in captivity and also display striking variation in behavioral traits, making them good model organisms for examining the relationship between innovation and other behavioral traits. We measured persistence, motor diversity, motivation, activity, efficiency, inhibitory control, and neophobia demonstrated by hyenas while interacting with the MAB. We also independently assessed inhibitory control with a detour cylinder task. Most hyenas were able to solve the MAB at least once, but only four hyenas satisfied learning criteria for all four possible solutions. Interestingly, neither measure of inhibitory control predicted repeated innovation. Instead, repeated innovation was predicted by a proactive syndrome of behavioral traits that included high persistence, high motor diversity, high activity and low neophobia. Our results suggest that this proactive behavioral syndrome may be more important than inhibitory control for successful innovation with the MAB by members of this species.

INTRODUCTION

Innovation is the ability to invent novel solutions to existing problems and solve novel problems (Reader et al., 2016). A growing body of research has shown that innovation is adaptive in a wide array of species for invading novel environments (Lefebvre, Reader, & Sol, 2004). Innovation in the wild is related to brain size, behavioral flexibility, general intelligence, culture, and even anatomical evolution and speciation (for review see Reader et al. 2016). Spontaneous innovation in the wild can be very difficult to observe in some species. For this reason, many researchers have begun to use experimental problem-solving paradigms to measure innovative abilities in both captive and wild animals (Auersperg, von Bayern, Gajdon, Huber, & Kacelnik, 2011; Benson-Amram, Weldele, & Holekamp, 2013; Borrego & Dowling, 2016). Innovative problem-solving paradigms typically require a subject to perform a specific motor pattern to obtain food from an apparatus and are sometimes referred to as 'extractive foraging tasks'. However, experimental assessments of innovation via problem-solving paradigms can be problematic because many traits, some of which are "non-cognitive", affect problem-solving, and variation in any of these traits might explain variation in success. Although the traits that underlie problem-solving are also those predicted to underlie innovation (Griffin & Guez, 2014), the influence of non-cognitive traits calls into question the validity of problem-solving as a measure of cognition. For example, a recent study suggests that problem-solving ability in dogs may not actually be a separable construct from temperament (Bray, Sammel, Seyfarth, Serpell, & Cheney, 2017). To further investigate the relationship between problem-solving and cognition, we aimed to concurrently examine several traits, including inhibitory control, that may underlie problem-solving using a multi-access box paradigm that requires repeated innovation and learning.

Inhibitory control is one important cognitive trait thought to underlie innovation and problem-solving; it is the ability to resist performing a prepotent or previously learned behavior when it is inappropriate, such that the behavior will yield no reward (MacLean et al., 2014). Inhibitory control is widely believed to be an important component of innovative problem-solving because individuals must inhibit previously learned responses and innate biases in order to develop a novel behavioral solution to a familiar problem (Manrique, Völter, & Call, 2013). Inhibitory control is well-established as an important executive function in human cognition (e.g. Mischel et al. 1989), and it is also a well-studied aspect of animal cognition (Kabadayi et al., 2018). Overall, problem-solving and inhibitory control are often discussed together, with failures in problem-solving often attributed to inadequate inhibitory control (Manrique et al., 2013; Taylor, Hunt, Medina, & Gray, 2009; Alex Thornton & Samson, 2012). Although inhibitory control is strongly correlated with success in various cognitive tests among primates (Burkart et al., 2017; Herrmann et al., 2007), two studies that have independently measured both inhibitory control and problem-solving in other animals failed to find straightforward relationships between them. First, Shaw (2017) found no correlation between performance on an inhibitory control task and a problem-solving task in wild North Island robins. Second, problem-solving performance by domestic dogs was related to their performance in two inhibitory control tasks, but the relationship was negative for one task and positive for another (Müller et al. 2016). Thus, the relationship between an individual's level of inhibitory control and its problem-solving ability remains unclear.

Other non-cognitive behavioral traits that have been measured in relation to problem-solving success include persistence, motivation, motor diversity, neophobia, efficiency and activity (Chow, Lea, & Leaver, 2016; Griffin & Guez, 2014; Sih & Del Giudice, 2012).

Persistence typically correlates positively with problem-solving success; individuals (and perhaps also species) that engage more with tasks are more likely to solve them (Benson-Amram et al. 2016; Griffin and Guez 2014). Interestingly, motivation does not typically seem to correlate with problem-solving success separately from its influence on persistence (Griffin and Guez 2014). Motor diversity also has a strong positive relationship with problem-solving success; the number of different motor actions that an animal uses predicts success (Griffin and Guez 2014). Low neophobia sometimes predicts success between individuals, though evidence on this is mixed (Griffin and Guez 2014). Efficiency, or the amount of time taken to solve a task, is often used as the dependent measure in problem-solving studies (Chow et al., 2016). However, individuals that are faster at solving novel problems are sometimes less accurate across trials or tasks (Sih & Del Giudice, 2012). Activity, the general physical activity level of an individual, is thought to be related to high efficiency in problem-solving, but also less flexibility in relation to tasks like reversal learning (Brust, Wuerz, & Krüger, 2013; Schuster, Zimmermann, Hauer, & Foerster, 2017; Sih & Del Giudice, 2012).

The multi-access box (MAB) paradigm is ideal for addressing questions about the relationship between cognitive and non-cognitive behavioral traits in problem-solving studies. A MAB is a problem-solving apparatus that offers a novel way of assessing innovation. A MAB has multiple, unique entry points or solutions on a box to a common interior that is baited with food rewards (Auersperg et al., 2011; Huebner & Fichtel, 2015; Manrique et al., 2013). Different MAB solutions may require sequential learning of skills needed to open the box (Huebner & Fichtel, 2015), or they may require different cognitive skills (Auersperg et al., 2011) or motor actions (Manrique et al., 2013). A MAB can be used to measure both innovation and learning because subjects can be scored on several variables across successive trials. For example, innovation can be measured as finding a novel solution and learning can be measured as the number of solutions a subject learns to open reliably across multiple trials. Once a subject has learned one solution, it can be blocked so that the subject is required to learn a new solution to access further food rewards. Because subjects must inhibit using blocked solutions in order to learn new ones, the MAB paradigm allows researchers to directly measure inhibitory control by recording the amount of time spent on blocked solutions before discovering a new solution. As in traditional problem-solving tasks, other behavioral traits can also be assessed for their relative influence on performance with the MAB.

Here, we used a MAB (Auersperg et al., 2011) that we designed for use with mammalian carnivores. Our MAB had four solutions, each requiring a different motor action, all of which occur within the repertoires of most carnivore species. Repeated innovation was assessed as the number of these solutions learned. We use the term ‘repeated innovation’ because, unlike binary measures of innovation, achieving a high score requires subjects to innovate multiple times and demonstrate learning through repeated use of each solution. We also used the MAB to measure persistence, motor diversity, motivation, activity, efficiency, neophobia, and inhibitory control.

In addition to measuring inhibitory control exhibited by subjects while interacting with the MAB, we also used a standard ‘cylinder task’ to independently measure inhibitory control in the same subjects. The cylinder task is a detour-reaching task that requires an individual to inhibit the impulse to go straight toward a food reward easily visible inside a clear cylinder, and instead detour to the opening on either end of the cylinder to retrieve the food (Kabadayi et al., 2018). Inhibitory control on the cylinder task requires subjects to use a previously learned detour response while inhibiting an impulse to reach straight for food. By contrast, inhibitory control with the MAB requires inhibiting a previously learned response to a blocked solution and trying a novel behavior. Thus, these may represent different types of inhibitory control, each bearing a unique relationship to innovation. While some previous research has shown that both types of inhibitory control positively correlate with one another (MacLean et al., 2014), other studies suggest that there may be multiple types of inhibitory control that do not necessarily correlate (Brucks, Marshall-Pescini, Wallis, Huber, & Range, 2017).

We chose to assess repeated innovation with the MAB in captive spotted hyenas (*Crocuta crocuta*) because these animals have been established as good model organisms for testing hypotheses suggested to explain the evolution of intelligence (Holekamp et al., 2007); innovation has been previously assessed both in wild and captive subjects using a single-access puzzle-box (Benson-Amram et al., 2013). Captive hyenas readily participate in cognitive tests and they are fairly innovative relative to other carnivores (Benson-Amram et al., 2016). In addition, personality traits have been assessed in both captive (Gosling, 1998) and wild hyenas (Yoshida, Van Meter, & Holekamp, 2016). The striking variation among hyenas in regard to personality traits makes them a good model for understanding the role that personality and other non-cognitive behavioral traits might play in relation to repeated innovation. Finally, problem-solving success is correlated with brain size in captive carnivores, including spotted hyenas, suggesting that problem-solving may be an adequate paradigm for assessing innovative problem-solving in these animals (Benson-Amram et al., 2016).

Our first goal was to investigate the relationship between repeated innovation and inhibitory control. Based on previous research (MacLean et al., 2014), we expected that our measures of inhibitory control from the MAB and the cylinder task would correlate with one another. In addition, based on the suggested link between inhibitory control and failures in problem solving (Manrique et al., 2013), we predicted that both our measures of inhibitory control would be positively related to repeated innovation. Our second goal was to investigate the other behavioral traits associated with innovation and problem-solving, including persistence, motivation, motor diversity, activity, efficiency, and neophobia. In order to address the issue regarding whether successful problem-solving is affected by these 'non-cognitive' traits, we chose to use a MAB. The MAB paradigm was designed to feature a battery of tasks based on the idea that performance on any single task is unlikely to be fully representative of an individual's cognitive abilities (Auersperg et al., 2011). While performance on each task individually is likely to be affected by various traits, performance across different tasks should be a strong indication of cognitive ability (Burkart et al., 2017). Here, subjects were required to innovate multiple times using a diverse array of motor actions and demonstrate learning of each innovation. Because repeated innovation scores on our MAB may be more indicative of cognitive ability than success on a single task, we predicted that our non-cognitive traits, persistence, motivation, motor diversity, and neophobia, would have no relationship to repeated innovation. Our only exceptions from this prediction was with respect to efficiency and activity, both of which are associated with reduced accuracy or flexibility. We predicted that both of these would have a negative relationship with repeated innovation, which requires both accuracy and flexibility to achieve a high score.

METHODS

Subjects

We tested 10 captive spotted hyenas housed at two different institutions, 5 at the Oak Creek Zoological Conservatory (OCZC) (Madisonville, TX) and 5 at the Denver Zoo (Denver, CO). Subjects included 4 adult females, 3 adult males, 2 subadult females, and 1 subadult male.

All subjects had been located at their institutions for at least a year before testing began with the exception of “Wibari”, who had only been at OCZC for 2 months.

Apparatus

Our MAB was a square metal box 40.64 x 40.64 x 40.64 cm (length x width x height), weighing approximately 18 kg (Figure 2.1). The interior of the MAB was baited with a food reward. Each vertical side of the MAB provided access to the interior of the box via a solution that required a different motor behavior to open.

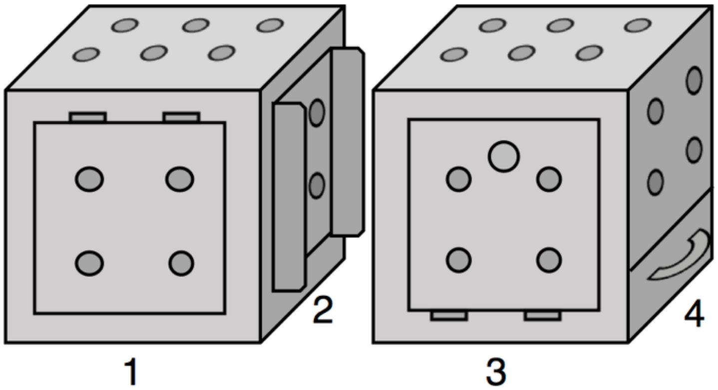

Figure 2.1. MAB used in the current study. **1)** The push flap solution; **2)** the sliding door solution; **3)** the pull flap solution; and **4)** the drawer solution. Small filled gray circles indicate the approximate number and location of holes drilled through the wall of the MAB. Large gray circle on side 3 represents the location of the door knob. Small rectangles represent the location of hinges.

The solutions were as follows: 1) The push flap: this was a door 30.5 x 28 cm with a hinge on the top that could be pushed inwards to open. 2) the sliding door: this was a door 30.5 x 28 cm with protruding flanges that could be pushed or pulled sideways to slide open. 3) The pull flap: this was a door 30.5 x 28 cm with a hinge on the bottom that could be pulled outwards and downwards to open by grasping a doorknob near the top of the door. 4) The drawer: this was a drawer 10 cm in height that was flush against the bottom of the MAB and took up the entirety of the floor of the MAB, it could be pulled outwards to open. Magnets were used to create mild resistance on all solutions to ensure that they would not accidentally fall open; subjects were required to actively interact with the MAB to retrieve the food. All solutions could be accessed using either mouth or paws. The top of the MAB was removable for familiarization trials. The MAB had multiple 2.5 cm circular holes cut on every side except the bottom so that subjects could smell the food inside during trials. All four solutions could be blocked by bolting them shut such that blocked solutions could still be manipulated by subjects but would not open to allow food retrieval.

In the cylinder task, we used two hollow cylinders. In familiarization trials we used an opaque cylinder made from PVC. In test trials we used a transparent cylinder made from cast acrylic plastic. Both cylinders measured 46 cm in length and 30.5 cm in diameter with a wall thickness of 1 cm. During testing, we placed a cylinder horizontally on the ground with both ends open for the retrieval of food through either end. Unlike previous studies that have used an inhibitory control cylinder, these cylinders were drilled with holes 4 cm in diameter evenly spaced across the wall of the cylinder. Spotted hyenas are highly sensitive to olfactory cues and this was done to ensure that the scent of the food would not spontaneously lead subjects to the correct behavior before they had seen the food through the wall of the transparent cylinder.

Procedure

Each subject was tested alone in its home enclosure. Subjects were not fasted prior to testing; instead testing took place during either their morning (AM) or evening (PM) feeding times, and food rewards were part of their normal daily diet, which included, eggs, ground meat, and pieces of meat and bone. The size of the food reward was kept to roughly 200 g. Each subject was temporarily moved to an adjacent enclosure and the MAB or the cylinder was then placed between 1-2 m from the entrance of the test enclosure, in a location clearly visible to both human observers and the subject when it re-entered the test enclosure. All trials were videotaped. Trials began when the subject entered the test enclosure and ended upon successful retrieval of the food or after 15 minutes had elapsed, whichever came first.

Subjects were first given 2 – 5 familiarization trials with the MAB, during which all four solutions were accessible and the top of the box was removed. All solutions were left accessible, rather than blocked, during familiarization trials to prevent subjects from wrongly learning that a solution could not be opened if they attempted to do so, but no hyenas retrieved food via any solution during familiarization trials. Familiarization trials ended when the individual retrieved the food or 15 minutes elapsed, whichever came first. Subjects were required to successfully retrieve food from the MAB in under 3 minutes on at least 2 consecutive trials to progress to the first test phase.

There were four phases of MAB testing. During the initial phase, the top of the box was in place and all four solutions were accessible. Once a subject used the same solution in 3 out of 4 consecutive trials, that solution was considered learned and the subject would progress to the next phase of testing. In phase 2, the learned solution from phase 1 was made inaccessible by screws that held it closed. In phase 3, the two previously learned solutions were inaccessible. In the final phase of testing all solutions but one were inaccessible. Subjects could fail to progress to subsequent phases in two ways: either they timed out (15 minutes elapsed without successful food retrieval) in 4 consecutive trials across more than 1 testing session or they required more than 7 trials to reach learning criterion, excluding trials in which the subject timed out. Seven trials were used as the cut-off for learning because the chance of reaching learning criterion by chance alone exceeded 50% at 8 trials during phase 1. We continued to use this criterion in subsequent phases largely because reducing it further for each phase would have interfered with our learning procedure that allowed for at least 4 trials per subject per phase.

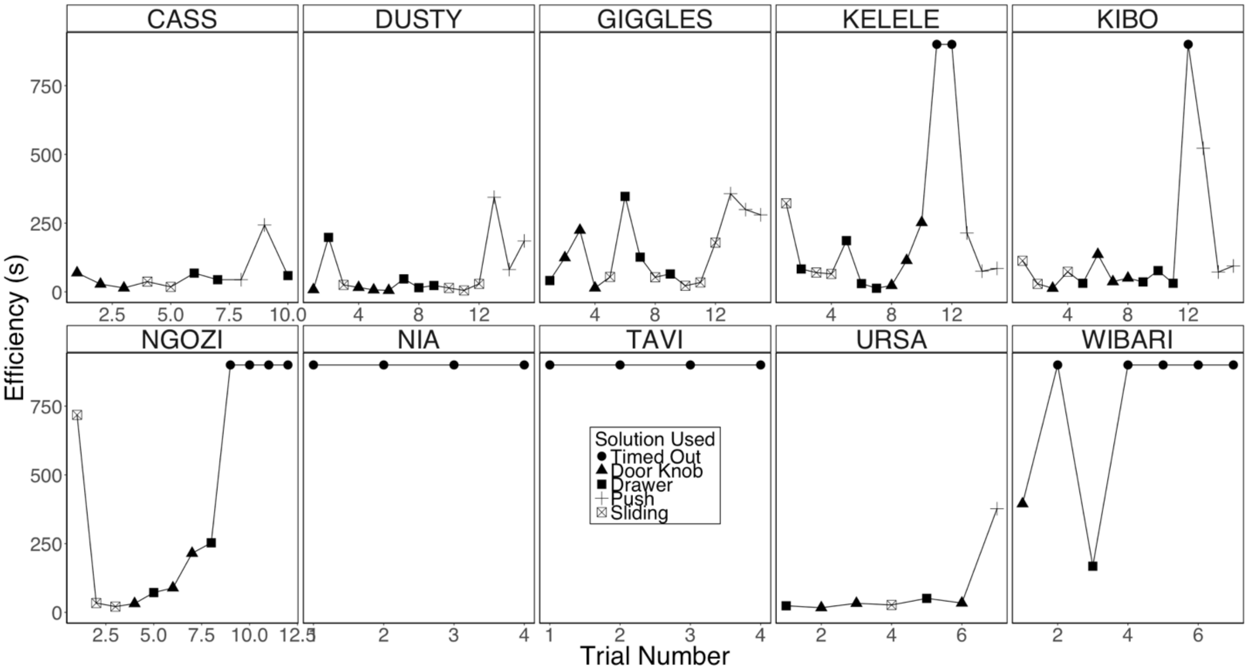

Figure 2.2. Problem-solving efficiency across trials for each subject. Note that the x-axis scales vary among subjects based on the number of trials that each subject received. Efficiency was measured as the number of seconds from first contact to food retrieval. In trials where subjects failed to retrieve food a score of 900 s was applied.

Ultimately, only 1 subject reached the 7-trial limit on phase 2, 1 subject required 5 trials for phase 2, and all other subjects reached criterion in 3 or 4 trials on phase 2, rendering an even stricter criterion unnecessary (Figure 2.2). Subjects were given a score of 0 through 4 for the number of solutions learned; this represented our measure of repeated innovation.

Before using the clear cylinder to test inhibitory control (hereafter inhibition-C), we gave subjects a minimum of 5 familiarization trials with the opaque cylinder. Subjects were required to pass at least 4 of 5 trials in order to proceed to test trials with the transparent cylinder. A ‘pass’ was defined as successful retrieval of the food without touching the outside of the cylinder. Subjects were allowed to retrieve the food regardless of whether they passed or failed. After familiarization trials, each subject was given 10 test trials with the transparent cylinder. Inhibition-C was scored as the number of passes a subject achieved out of 10 trials. A high inhibition-C score indicated good inhibitory control. The cylinder was always positioned perpendicular to the exhibit entrance to encourage an approach towards the long side of the cylinder such that subjects were actively required to detour to either side of the cylinder to obtain the food reward.

All subjects at the Denver Zoo were presented with the cylinder task first and all subjects at the OCZC were presented with the MAB first. On the first day of testing subjects were given three trials with each apparatus. On subsequent days subjects received two to five trials with each apparatus, with a maximum of six trials total involving both apparatuses per day. Subjects were tested approximately six days each week until completion of testing. Testing for all 10 subjects was completed over the course of two-week periods at both institutions.

Behavioral assays

All trial videos were coded using BORIS v. 2.97 (Friard & Gamba, 2016). Repeated innovation was measured as the number of the four possible MAB solutions learned by each subject (Table 2.1, #9). We defined learning as solving the MAB with the same solution in three of four consecutive trials. Successfully solving the MAB was defined as a trial in which the subject opened the box and retrieved the bait using one of the four possible solutions. The learning criterion was based on preliminary observations of carnivores interacting with the MAB. We verified this criterion by recording the amount of time subjects spent in contact with the solution that they ultimately used to retrieve food from the MAB (work time, Table 2.1, #7), and inquiring how work time changed across successful trials.

Table 2.1. Summary of behavioral measures used in analysis.

Term	Definition
Measures extracted once per trial	
1. Persistence	Ratio of time spent contacting the MAB to total trial time
2. Motivation	Latency from trial start to first contact with the MAB
3. Motor diversity	# of 5 behavior patterns used while contacting the box
4. Activity	Ratio of time spent contacting the box through biting or pawing to total time spent in contact
5. Efficiency	Latency from first contact to food retrieval; unsuccessful trials did not receive a work time score
6. Inhibition-M	Amount of time spent in contact with blocked solutions relative to total contact time
7. Work Time*	Duration of time subjects spent in contact with the solution that they ultimately used to retrieve food from the MAB
Measures extracted once per subject	
8. Neophobia	Latency from trial start to food retrieval on first familiarization trial
9. Repeated innovation	score of 0 -4 indicating the # of MAB solutions learned
10. Inhibition-C	score of 0 – 10 with the inhibitory control cylinder representing the number of successful trials

Note: Measures #1-6 were extracted from each individual trial resulting in multiple measures per subject. Measures #7-9 were taken once per subject.
*Work time was only used to assess learning and was not compared to repeated innovation.

Six specific measures were calculated once for each trial (Table 2.1, #1-6). Persistence is typically described as the extent to which subjects engage with an apparatus, usually measured as contact time, work time, or number of attempts (Griffin & Guez, 2014; van Horik, Langley, Whiteside, & Madden, 2017). Here, persistence was extracted as the ratio of time spent in contact with the MAB to the total duration of the trial. Contact was defined as touching the MAB with the snout or paws. Scores close to 1 indicated high persistence and scores close to 0 indicated low persistence. We specifically used a ratio to measure persistence, instead of total time in contact with the MAB, to avoid confounding persistence with efficiency.

Motivation, the degree to which an animal wants to engage with a task (typically because of desire for food rewards), is closely related to persistence in that persistence can be considered a measure of motivation to work on a task. Motivation is typically measured as either body condition, some other measure of hunger, or latency to approach an apparatus (Griffin & Guez, 2014; van Horik et al., 2017). Here, motivation was extracted as the latency from trial start to first contact. A high score here indicated a slow approach and low motivation whereas a low score indicated a fast approach and high motivation. Although persistence can sometimes be evidence of motivation, measures of motivation do not always correlate with measures of persistence. Furthermore, unlike persistence, motivation is not always linked with problem-solving success (Griffin & Guez, 2014). Because persistence and motivation appear to be distinct (though related) constructs, we chose to use measures of both persistence and motivation here. Because we measured persistence as a ratio of time spent in contact, and motivation as latency to first contact, our measures of motivation and persistence did not overlap, so any correlation might be due to persistence and motivation deriving from a shared latent factor (e.g., hunger) rather than confounding of measurement.

Motor diversity is the total number of unique motor patterns that a subject may exhibit in a problem-solving context (Diquelou, Griffin, & Sol, 2015; Griffin & Guez, 2014). This has also been referred to as behavioral variety (Chow et al., 2016), exploration diversity (Benson-Amram et al., 2013), and behavioral diversity (Benson-Amram et al., 2016). Here, motor diversity was extracted by calculating a score of 0-5 for the number of different motor patterns used to contact the MAB. The 5 motor patterns were a) sniffing or contacting the MAB with the snout (but not biting), b) biting the MAB, c) flipping the MAB, d) using one or both paws to dig at the MAB, and e) moving the MAB by pushing, dragging or carrying. A score of 5 indicated high motor diversity and a score of 0 indicated low motor diversity.

Activity, or the physical activity level of a subject (Brust et al., 2013; Sih & Del Giudice, 2012) was measured in the context of exploration with the MAB. We extracted activity as the ratio of time spent using contact behaviors that involved the use of mouth or paw to the total time spent in contact (i.e. the ratio of motor patterns b-e to patterns a-e). Activity therefore referred to using behaviors, during contact with the MAB, that had the potential to lead to solutions. A high score indicated highly active exploration and a low score indicated more passive exploration of the MAB.

We used the term efficiency to mean the latency to problem-solving success in each trial (after Chow et al. 2016). Although this measure is often sometimes used as the dependent variable to measure success in many problem-solving studies (e.g. Chow et al. 2016), here, we analyzed it as an independent variable. Efficiency was extracted as the latency from first contact with the MAB to food retrieval (i.e. the amount of time it took to solve the MAB). Subjects who were not successful at retrieving food did not receive an efficiency score in that trial. A low score here indicated quick retrieval of the of the food and high efficiency, whereas a high score indicated slow retrieval and low efficiency. Inhibitory control as measured in MAB trials (hereafter inhibition-M) was extracted as the amount of contact time spent on a blocked solutions relative to total contact time. A low inhibition-M score indicated good inhibitory control.

To further assess that these trial measures were independent we examined the one-to-one correlation between all trial-level measures (Table 2.2). Only persistence and motivation were correlated at $P < 0.05$ (Spearman's $\rho = -0.51$), which was not unexpected and likely due to a shared latent factor (see above).

Table 2.2. Correlation matrix of six raw per trial behavioral measures from Table 2.1.

	1	**2**	**3**	**4**	**5**	**6**
1. Persistence	1.00	-0.48*	0.12	-0.16	-0.11	-0.13
2. Motivation		1.00	-0.02	0.21	0.07	0.27
3. Motor diversity			1.00	-0.01	0.47	0.24
4. Activity				1.00	-0.14	0.11
5. Efficiency					1.00	0.18
6. Inhibition-M						1.00

*Indicates statistical significance at $\alpha = 0.05$, Spearman's rank correlation.

In addition to repeated innovation and inhibition-C, we extracted one more measure once per subject. Neophobia is typically described as aversion to novelty, normally measured as latency to approach novel objects or food (Greggor, Thornton, & Clayton, 2015). Here, neophobia was extracted on each subject's first familiarization trial as the latency from the start of the trial to food retrieval. This is distinct from motivation, which was measured on each test trial (but not on familiarization trials) as latency from the start of the trial to first contact, rather than food retrieval. Neophobia has previously been measured in spotted hyenas as latency to contact a novel object with or without bait (Benson-Amram et al., 2013; Greenberg & Holekamp, 2017). Elsewhere, neophobia is often measured as latency to feed near a novel object (Greggor et al., 2015), and in the current study we measured neophobia as latency to feed from inside a novel object by measuring each subject's latency to feed from the MAB the first time it encountered the MAB during its initial familiarization trial (Table 2.1, #8). We thought this measure best assessed neophobia because most hesitancy our subjects displayed towards the MAB was in regard to putting their heads inside the MAB to feed. Here a high score indicated a slow approach towards the MAB and high neophobia, whereas a low score indicated a fast approach and low neophobia.

Several of these behavioral traits are often considered "personality traits" when they are repeatable across both time and context. Here, most of our measures were taken across multiple trials, which allowed us to test repeatability, but we only measured them in the context of the MAB. However, two traits, neophobia and inhibition-C, were only scored once. Ultimately, there were two categories of variables used in the final analysis: variables that were measured per trial while subjects interacted with the MAB (Table 2.1, # 1-7), and variables measured once per subject (Table 2.1, #8-10).

Reliability

All videos were coded by LJU. A random subsample of 20% of videos were coded by an independent coder for reliabilities on durations and frequencies of raw behaviors (Pearson's correlation coefficient: $r = 0.99$). A separate random subsample of 20% of videos were coded by another independent coder for latency to first contact (Pearson's correlation coefficient: $r = 0.98$) and latency to food retrieval (Pearson's correlation coefficient: $r = 0.99$). Inter-observer reliability assessments were conducted on the raw durations, latencies, and frequencies of behaviors.

Statistical analysis

To validate our learning criterion, we used a linear mixed model (LMM) to examine how work time changed across trials. Work time was used as the dependent variable and we included trial number for the solution used as a fixed effect and subject ID as a random effect in the model.

To examine the traits (#1-6, 8, and 10 in Table 2.1) influencing the number of MAB solutions learned we used a generalized linear model (GLM) with a Poisson error and log link. To avoid over-parameterization of our model, we systematically pared down the number of predictor variables in four steps. First, we used *t*-tests for unequal variance to check for any differences in the number of solutions learned based on zoo, sex, and age class. Where no differences existed, we pooled the data in our model for predicting the number of MAB solutions learned. Otherwise, we included them in our model to control for their effect.

Second, because number of MAB solutions learned was a subject-level variable we converted the repeated trial-by-trial measures (#1-6, Table 2.1) to mean scores for each subject. Prior to obtaining means, to ensure that means were valid representations of each subject's behavior, we looked for individual differences and repeatability within individuals in our trial measures. To do so, we used a likelihood ratio test to compare linear mixed models (LMMs) with and without subject ID as a random effect. Each of these six (#1-6, Table 2.1) variables was used as a response variable in its own LMM, and as predictor variables we included age class, sex, AM or PM, zoo, phase, phase trial number, total trial number, and an interaction between phase and phase trial number to control for their effects. From the LMMs we also calculated intraclass-correlation coefficients (ICCs) to examine repeatability. For those measures in which ID was both significant and repeatable, we took a mean score for each subject for eventual comparison to the number of solutions learned. Any measures in which ID was not significant in the LMM were excluded from the final model because this indicated to us that variation across trials within subjects was larger than variation between subjects, and therefore that a mean score would not be a valid representation of an individual subject's performance.

Third, we checked for any correlations between independent measures to avoid multicollinearity when predicting the number of MAB solutions learned. As it is not recommended to include variables that correlate at higher than 0.70 (Tabachnick & Fidell, 2013), where correlation coefficients exceeded 0.70, we used principal component analysis to extract a composite score, representing a syndrome of non-cognitive behaviors. Ultimately only two variables were included in our final model and we conducted no more paring down after this step. The 'rule of 10' suggests regression models should not include more than one variable per ten subjects (Peduzzi, Concato, Kemper, Holford, & Feinstein, 1996). However, review of this rule with logistic models suggested that, although type II error increases with fewer than ten subjects per factor, type I error does not increase substantially. Therefore, null results must be interpreted cautiously, but significant results can be interpreted normally, albeit with a degree of caution regarding generality (Vittinghoff & McCulloch, 2007). All statistics were run using R 3.2.2 GUI 1.66 Mavericks build (6996) (Bates et al. 2015; Kuznetsova 2017; Lüdecke 2018; R Core Team, 2016; Wickham et al. 2017; Wickham 2009).

RESULTS

Of the ten hyenas tested, eight opened the box at least once, six learned at least 1 solution and four learned all 4 solutions on the MAB (Table 2.3). In sum, hyenas participated in a total of 104 trials (mean trials total per subject ± SD = 10.40 ± 4.62, range = 4 – 15 trials per subject). The push solution consistently appeared to be the last solution learned, suggesting that it was harder than the other solutions. However, there was no consistent pattern to the order in which other solutions were used (Figure 2.2). Inhibition-C scores varied from 3 to 9 successful trials of 10 total trials (Table 2.3).

Table 2.3. Scores for each subject on neophobia, inhibition-C, and the number of solutions learned on the MAB.

ID	MAB	Inhibition-C	Neophobia (s)	Age Class	Sex	Institution
Cass (C)	1	5	13	Adult	Female	OCZC
Dusty (D)	4	9	9	Adult	Male	OCZC
Giggles (G)	4	8	300	Adult	Female	OCZC
Kelele (KL)	4	8	6	Subadult	Male	Denver
Kibo (KB)	4	4	33	Adult	Male	Denver
Ngozi (NG)	2	8	14	Adult	Female	Denver
Nia (NI)	0	6	900	Subadult	Female	Denver
Tavi (T)	0	9	669	Subadult	Female	Denver
Ursa (U)	0	5	17	Adult	Female	OCZC
Wibari (W)	0	3	372	Adult	Male	OCZC

Learning

To assess learning we used linear mixed models to examine how work time changed across trials. Work time significantly decreased across successful trials (LMM: $t_{78.34} = -2.12$, $P = 0.04$, n = 84 successful trials, N = 8 subjects; Figure 2.3). This showed that subjects became faster at solving the MAB with experience and suggested that our criterion of using the same solution on 3 out of 4 consecutive trials was sufficient for requiring that learning occurred.

Zoo, age class, and sex differences

The number of solutions learned did not differ between institutions (two-sample *t*-test assuming unequal variances: $t_8 = -0.16$, $P = 0.88$), age classes (two-sample *t*-test assuming unequal variances: $t_3 = 0.54$, $P = 0.63$) or the sexes (two-sample *t*-test assuming unequal variances: $t_6 = -1.53$, $P = 0.18$). Therefore, data from all ten hyenas were pooled into a single analysis.

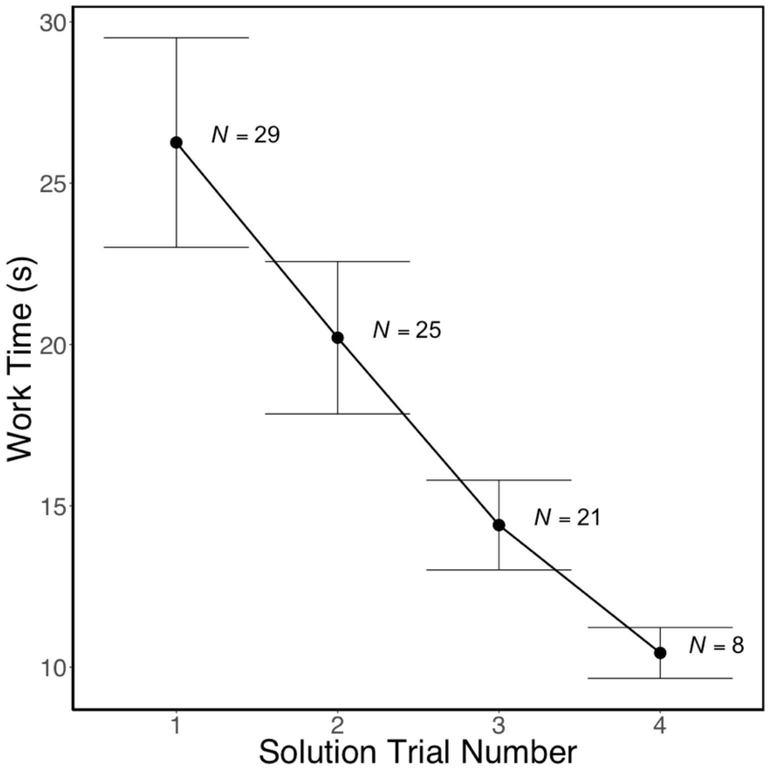

Figure 2.3. Learning curve across trials. Work time (s) was calculated as the amount of time subjects spent in contact with the solution that was ultimately used to retrieve food. Solution trial number corresponds to the number of trials with each particular solution. Error bars indicate standard error. N varies across trials because subjects opened solutions variable numbers of times.

Repeatability of measures

Because we were interested in subject level variation on each of our six trial measures for comparison to repeated innovation scores, prior to running any analyses we first checked for significant variation between subjects and repeatability within subjects on each measure using likelihood ratio tests (N = 58 to 104 trials). Of the 6 trial measures, only 3 showed significant variation among subjects across trials (Figure 2.4): persistence (Likelihood ratio test: $\chi^2_{(1, N = 104)} = 31.49$, $P < 0.001$, ICC = 0.60), motor diversity (Likelihood ratio test: $\chi^2_{(1, N = 104)} = 8.22$, $P < 0.01$, ICC=0.21), and activity (Likelihood ratio-test: $\chi^2_{(1, N = 104)} = 22.46$, $P < 0.001$, ICC = 0.40).

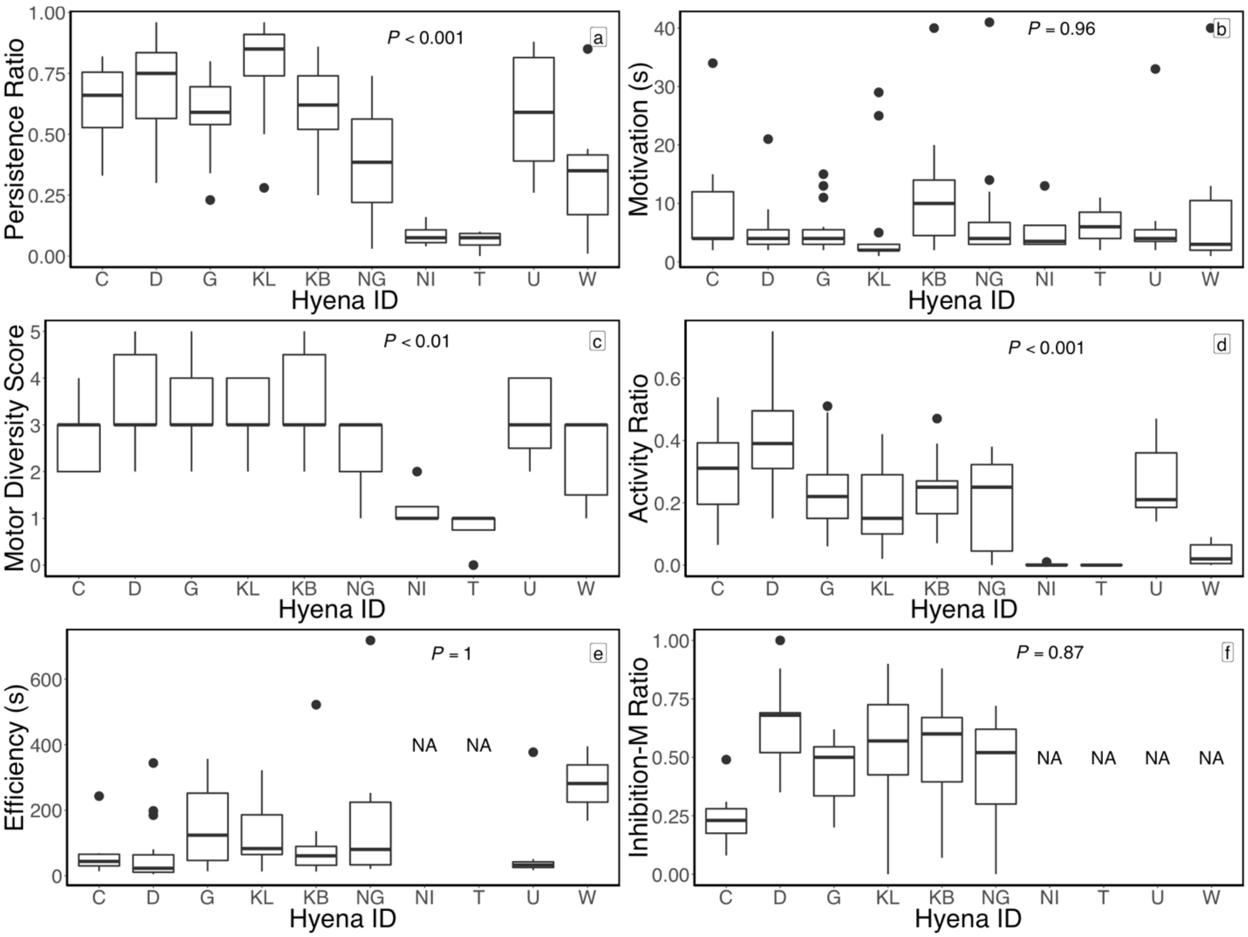

Figure 2.4. Variation among subjects on the six trial measures (#1–6, Table 2.1). Box and whisker plots show mean (± SD) scores for measures #1–6 in Table 2.1. Black points represent outliers. *P*-values are from likelihood ratio tests comparing models with and without subject ID while including control variables. Two subjects, NI and T, were not assigned scores (NAs) for work time because hyenas had to have at least one successful trial in order to achieve a work time score. Four subjects, NI, T, W, and U were not assigned scores for inhibition-M because inhibition-M scores could only be assigned for trials in phases 2–4 where at least one solution was blocked.

Motivation (Likelihood ratio test: $\chi^2_{(1, N = 101)} = 0$, $P = 0.96$, ICC = 0.002), efficiency (Likelihood ratio test: $\chi^2_{(1, N = 84),} = 0$, $P = 1$, ICC = 0.00), and inhibition-M (Likelihood ratio test: $\chi^2_{(1, N = 58)} = 0.03$, $P = 0.87$, ICC = 0.00) did not show significant variation among subjects across trials (Figure 2.4). This lack of significance, in addition to extremely low ICCs, indicates high variation and low repeatability within subjects on these variables, so they were excluded from further analysis. Mean scores for each subject on persistence, motor diversity, and activity were obtained for the next steps of analysis (Table 2.A.1).

Independence of measures

Next, we checked for correlations between mean persistence, mean motor diversity, mean activity, neophobia and inhibition-C prior to including them in the model as predictors of the repeated innovation score. Mean persistence, mean motor diversity, mean activity, and neophobia, but not inhibition-C, were correlated at higher than 0.70 (Table 2.4).

Table 2.4. Correlation matrix among subjects' scores on behavioral measures from Table 2.1.

	1	**3**	**4**	**8**	**10**
1. Mean persistence	1.00	*0.85	*0.81	*-0.90	0.02
3. Mean motor diversity		1.00	*0.78	*-0.67	-0.08
4. Mean activity			1.00	*-0.70	-0.08
8. Neophobia				1.00	-0.20
10. Inhibition-C					1.00

Note: Table 2.4 differs from Table 2.2 in that it shows the correlations between measures at the subject level with means, rather than the trial level with raw scores.
*Indicates statistical significance at $\alpha = 0.05$, Spearman's rank correlation.

Therefore, we applied principal component analysis to mean persistence, mean motor diversity, mean activity, and neophobia scores. Both the Kaiser criterion (Kaiser, 1960) and scree test (Cattell, 1966) indicated that only the first component should be retained. The first component explained 90% of the variance and all four traits loaded equally onto it. Mean persistence, mean motor diversity, and mean activity all loaded moderately positively onto this component whereas neophobia loaded moderately negatively onto it (Table 2.A.2).

Factors predicting repeated innovation

We used each subject's score on the first component of our principal component analysis (hereafter proactivity scores) for inclusion in our final model in place of mean persistence, mean motor diversity, mean activity, or neophobia scores. We chose the term 'proactivity' because high scores on the first component closely matched the proactive end of the commonly studied proactive-reactive axis of behavior (Sih, Bell, Johnson, & Ziemba, 2004). With only two predictor variables in our final model, proactivity scores and inhibition-C, we conducted no further model reduction. Using this model, proactivity scores positively and significantly predicted repeated innovation scores (Poisson GLM: $z = 2.65$, $P = 0.02$; Figure 2.5; Table 2.A.3) whereas inhibition-C did not (Poisson GLM: $z = 1.54$, $P = 0.27$; Table 2.A.3). Note that, with a sample size of 10, null results should be interpreted cautiously as the risk of Type II error is high.

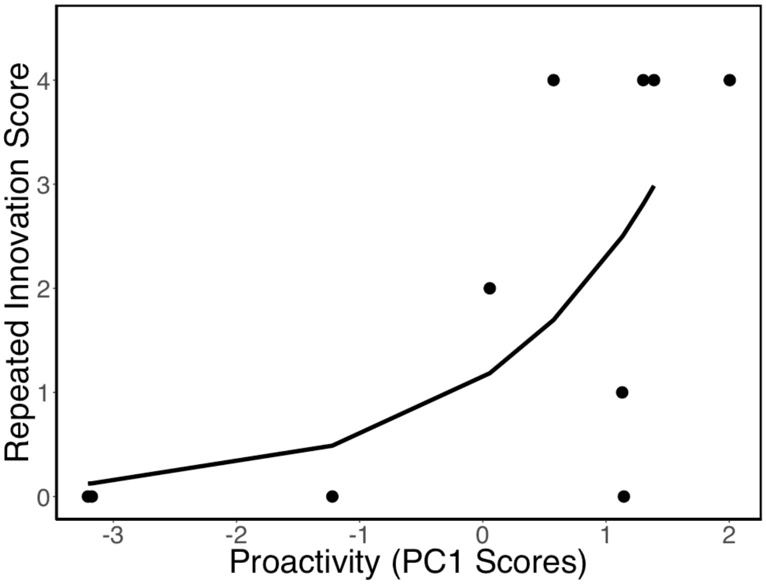

Figure 2.5. Relationship between PC1 scores and the number of MAB solutions learned (repeated innovation score). Each point represents one subject. The black line is a Poisson GLM regression of the number of MAB solutions learned against PC1 scores.

We had intended to also look at a correlation between inhibition-C and inhibition-M, but we were unable to do so because inhibition-M was measured multiple times whereas inhibition-C was only measured once. For comparison we could have taken a mean inhibition-M score; however, we did not calculate mean inhibition-M scores due to low repeatability within subjects.

DISCUSSION

We examined the effects of persistence, motivation, motor diversity, activity, efficiency, and two measures of inhibitory control on repeated innovation. We found no support for our hypothesis that inhibitory control would positively predict repeated innovation. Inhibition-C failed to predict repeated innovation, and inhibition-M did not vary significantly among hyenas. Opposite to our predictions, several of our non-cognitive traits strongly predicted repeated innovation through a composite proactivity score. High motor diversity, high persistence, high activity, and low neophobia were associated with higher repeated innovation scores. Not only had we predicted that most of these traits would have no relationship with repeated innovation, but we also predicted that activity would have a negative, not positive, relationship. Like inhibition-M, motivation and efficiency were not repeatable within subjects and therefore were not comparable to repeated innovation scores.

Inhibitory control and repeated innovation

The lack of a significant relationship between inhibitory control and repeated innovation was in contradiction to previous research on inhibitory control and problem-solving (Burkart et al., 2017; Herrmann et al., 2007; Manrique et al., 2013; Müller et al., 2016; Alex Thornton & Samson, 2012). Although we observed a great deal of variation in repeated innovation, inhibition-M was not repeatable within hyenas. This might suggest that inhibiting previous learning with the MAB did not affect the ability to learn a new solution. However, our sample size for inhibition-M was only 6 hyenas which may have been too small a sample in which to observe an effect. Inhibition-C was also not a significant predictor of MAB scores. The confidence interval for the estimated coefficient for inhibition-C overlapped zero which suggests the possibility that inhibition-C had no relationship with repeated innovation. On the other hand, with a sample size of only 10 subjects we also cannot rule out the high risk of type II error. Because of that risk, we also examined the effect size of the estimates for inhibition-C and proactivity. At the upper end of the confidence interval, an increase of 1 in inhibition-C scores corresponded with an increase of $e^{0.39} = 1.48$ in MAB scores, suggesting a potential relationship, but this effect size was quite small relative to the effect of proactivity in the GLM (Table 2.5). The few other studies that independently assessed individual's inhibitory control and problem-solving ability have also failed to find a simple positive relationship (Müller et al., 2016; Shaw, 2017). Although we had also planned to examine the relationship between inhibition-M and inhibition-C in this study, because inhibition-M was not repeatable within hyenas, we were unable to correlate inhibition-M and inhibition-C. Overall, our results suggest that variation among individuals on either type of inhibitory control may not be related to variation in repeated innovation; at best, the effect of the relationship is small relative to effects of other behavioral traits. Our results highlight the need for further study into how individuals use inhibitory control and how that contributes to successful innovation and learning.

Non-cognitive factors influencing repeated innovation

In the current study, higher repeated innovation scores were predicted by a composite proactivity score that consisted of high persistence, high activity, high motor diversity, and low neophobia. Previous research with hyenas also found consistent individual differences in motor diversity, and found that high motor diversity, high persistence and low neophobia predicted problem-solving success with a single-access puzzle box (Benson-Amram et al., 2013). Previous research with the MAB paradigm in keas and corvids produced results similar to those obtained in the current study. Birds that were less neophobic and more exploratory with the MAB were more likely to learn more solutions (Auersperg et al., 2011).

Not only were the results of the current study in contradiction to our predictions, but, persistence, activity, motor diversity, and neophobia were also both repeatable within subjects and strongly correlated with one another, suggesting that captive hyenas exhibit a “behavioral syndrome” within a problem-solving context. Sih et al. (2004) specifically define a behavioral syndrome as “a suite of correlated behaviors within a given behavioral context or across different contexts.” The current behavioral syndrome appears to closely mirror a syndrome of personality on the proactive-reactive axis (Sih et al., 2004). Proactive individuals are less neophobic, more persistent, active and exploratory, whereas reactive individuals show the opposite pattern (Sih et al., 2004). Here, proactivity was positively associated with repeated innovation; hyenas that learned more MAB solutions were more persistent, more active during exploration, had higher motor diversity scores and had lower neophobia scores.

A growing body of literature links behavioral syndromes and personality traits to cognition (Carere & Locurto, 2011; Griffin, Guillette, & Healy, 2015; Guillette, Naguib, & Griffin, 2017; Sih & Del Giudice, 2012). Current hypotheses that describe the adaptive function of proactive-reactive syndromes with regards to cognition do not suggest whether proactive or reactive individuals should be better at learning and problem-solving, but instead suggest that the proactive-reactive axis may reflect a trade-off between alternative learning styles in which proactive individuals are more innovative and faster at learning novel tasks whereas reactive individuals are slower and less innovative, but more flexible or accurate learners (Carere & Locurto, 2011; Ducatez, Audet, & Lefebvre, 2015; Griffin, Guez, Lermite, & Patience, 2013; Sih & Del Giudice, 2012). This trade-off is believed to have ecological relevance because it represents two alternative strategies for survival in the wild (Chittka, Skorupski, & Raine, 2009; Sih & Del Giudice, 2012). In the current study, no clear trade-off exists; proactive hyenas had higher repeated innovation scores and therefore also showed a high degree of flexibility by switching between four different solutions to the MAB. In addition, proactive hyenas were not more efficient at opening the MAB than reactive hyenas who learned fewer solutions. Indeed, whether proactive or reactive individuals perform better or worse on a task may be heavily task-dependent or may not be apparent in captivity where the costs and benefits of one strategy over another are reduced or absent. Other studies have also shown inconsistencies in the relationship between proactivity-reactivity and cognitive trade-offs (Bousquet, Petit, Arrivé, Robin, & Sueur, 2015; Schuster et al., 2017; Titulaer, van Oers, & Naguib, 2012). Of note, the proactive syndrome observed in our study was only measured in the context of the MAB which could account for the lack of a clear trade-off.

Conclusion

In sum, we found that a behavioral syndrome consisting of persistence, motor diversity, activity and neophobia predicted repeated innovation in a MAB paradigm with captive spotted hyenas. Contrary to our predictions, neither of our two measures of inhibitory control bore a clear relationship to repeated innovation scores. Although inhibitory control may be required for many problem-solving tasks, here the effect of inhibitory control was small or non-existent relative to the effect of proactivity scores. The finding that a proactive syndrome of correlated traits, including activity, motor diversity, persistence, and neophobia significantly predicted repeated innovation score was also contrary to our prediction that these traits would have either no effect or a negative effect. Overall, our results add support to the increasing body of literature that suggests that motor diversity, activity, persistence and neophobia have a strong relationship with innovative problem-solving and that these traits are often strongly interrelated. However, our results contradict hypotheses suggesting the direction of the relationship between proactive traits and innovation. Our results also shed some light on the murky relationship between behavioral traits, problem-solving, and cognition.

The relationship between behavioral traits and problem-solving is a burgeoning field (see Guillette & Griffin 2017), and though some consider these traits to be non-cognitive factors to be controlled when trying to measure variation in cognition (Rowe & Healy, 2014; van Horik & Madden, 2016), a relationship between behavioral traits and problem-solving does not necessarily negate cognitive explanations for problem-solving success (Griffin et al., 2015; Guillette et al., 2017). Here, repeated innovation required memory of previous solutions in order to reach learning criterion such that success could not be attributed solely to these non-cognitive behavioral traits. However, the strong influence of these non-cognitive behavioral traits does support the hypothesis that innovation and innovative problem-solving might not be exclusively cognitive constructs themselves. Although the proximate relationship between these traits and cognition remains unclear, ultimately, if an individual is highly innovative, it may be adaptive to perform behaviors that both encourage learning and increase opportunities for innovation, resulting in correlated evolution (e.g. Griffin 2016).

APPENDIX

Table 2.A.1. Means and standard deviations for each subject on each trial measure.

ID	**Persistence**		**Motivation**		**Motor Diversity**		**Activity**		**Efficiency**		**Inhibition-M**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**	**MD**	**SD**	**M**	**SD**
C	0.63	0.17	9.11	10.36	2.80	0.79	0.30	0.16	62.40	66.27	0.24	0.13
D	0.69	0.19	5.33	4.69	3.47	1.06	0.40	0.16	66.80	98.72	0.64	0.21
G	0.58	0.17	5.60	4.05	3.20	0.86	0.23	0.13	147.93	123.73	0.44	0.14
KL	0.79	0.19	5.86	9.05	3.33	0.62	0.20	0.12	117.85	95.96	0.55	0.26
KB	0.61	0.20	11.40	9.69	3.67	0.98	0.24	0.11	93.86	128.24	0.54	0.24
NG	0.38	0.24	8.42	10.89	2.42	0.79	0.19	0.15	179.25	234.35	0.45	0.25
NI	0.09	0.05	5.75	4.86	1.25	0.50	0.00	0.01	NA	NA	NA	NA
T	0.06	0.05	6.33	4.51	0.75	0.50	0.00	0.00	NA	NA	NA	NA
U	0.59	0.25	8.14	11.07	3.14	0.90	0.27	0.13	80.43	131.20	NA	NA
W	0.34	0.28	9.86	13.97	2.29	0.95	0.04	0.04	281.50	160.51	NA	NA

Table 2.A.2. Results of the principal components analysis.

	PC1	**PC2**	**PC3**	**PC4**
Proportion of variance	0.91	0.04	0.04	0.01
Factor loadings				
Mean persistence	-0.51	0.44	-0.19	0.72
Mean motor diversity	-0.51	0.38	-0.36	-0.68
Mean activity	-0.49	-0.82	-0.30	0.07
Neophobia	0.49	0.03	-0.86	0.11

Table 2.A.3. Parameter estimates from the Poisson GLM comparing variation in repeated innovation scores. N = 10 hyenas.

	Estimate	**SE**	**95% CI**	***z***	***P***
Intercept	-0.74	0.91	-	-0.81	0.42
Proactivity	0.61	0.26	0.10, 1.11	2.65	0.02*
Inhibition-C	0.14	0.13	-0.11, 0.39	1.54	0.27

*Indicates statistical significance at α = 0.05.

CHAPTER THREE

INNOVATIVE PROBLEM-SOLVING IN WILD HYENAS IS RELIABLE ACROSS TIME AND CONTEXT

Johnson-Ulrich, L., Holekamp, K. E., & Hambrick, D. Z. (submitted to *Scientific Reports*)

ABSTRACT

The question of how cognition evolves across the animal kingdom remains unresolved despite more than a century of intensive research. Recently, there has been growing recognition of the value of intraspecific studies of wild, free-ranging populations because they allow researchers to examine the causes of variation in an ecologically valid context and also to examine the fitness consequences of this variation. However, for the intraspecific approach in wild animals to be useful for testing evolutionary hypotheses, researchers must provide evidence that that measures of cognition reflect stable, general traits. Here, we investigated the overall reliability (reflecting stability) of innovative problem-solving in wild, free-ranging spotted hyenas (*Crocuta crocuta*) and asked whether estimates of reliability were affected by individual factors such as age-sex class, contextual factors such as task demands and socio-ecological variables, and by temporal factors including the number of times subjects were tested and the length of the interval between tests. Overall, we found remarkably high levels of reliability for innovation compared to other animal behaviors, suggesting that innovation is a stable, general trait in wild spotted hyenas.

INTRODUCTION

The question of how cognition evolves across the animal kingdom remains unresolved despite more than a century of intensive research. The most common approach to addressing this question has been to compare average levels of cognition across species (Benson-Amram et al., 2016; DeCasien, Williams, & Higham, 2017; Dunbar & Shultz, 2007; Fristoe et al., 2017; MacLean et al., 2014). In this *interspecific* approach, individual differences within species are treated as random error (or “noise”). Recently, there has been growing recognition of the value of using individual differences to test evolutionary hypotheses—the *intraspecific* approach (Ashton, Thornton, et al., 2018). Intraspecific studies of wild, free-ranging populations are especially valuable for understanding cognitive evolution, because individual variation is the raw material on which natural selection acts. Thus, researchers may examine the causes of variation in an ecologically valid context and also to examine the fitness consequences of this variation (Cauchoix, Hermer, Chaine, & Morand-Ferron, 2017; Pritchard, Hurly, Tello-Ramos, & Healy, 2016; A. Thornton et al., 2014). Despite this recognition, in the field of cognitive ecology there have been few attempts to empirically test the hypothesis that measures of cognition reflect stable, general traits [see 7,8].

The hypothesis that a cognitive measure reflects a stable, general trait predicts that the measure should have a high degree of reliability: an animal that receives a high score on the measure at one point in time and in one context should receive a high score at later points in times and in other contexts, and vice-versa for an animal that receives a low score on the measure. As a psychometric concept, *reliability* refers to the amount of error contained in a measure, as reflected in the stability of the measure across contexts and time. It is especially important to demonstrate reliability of measures reflecting animal cognition in the wild, because there are many potential sources of error, including both external factors (e.g., weather, presence of conspecifics) and internal factors (e.g., hunger, stress) (Morand-Ferron, Cole, & Quinn, 2015; Pritchard et al., 2016; van Horik & Madden, 2016).

In this study, we assessed reliability of *innovation* across time and context in wild spotted hyenas using a problem-solving paradigm. Defined as the ability to solve a novel problem or use a novel behavior to solve a familiar problem, innovation is among the most commonly measured cognitive abilities in non-human animals (Reader & Laland, 2003). Although there has been a great deal of interest in the relationship between innovation and variables such as brain size, ability to invade new habitat, and life history traits in a diverse range of taxa, formal attempts to evaluate the reliability of problem-solving paradigms used to measure innovation remain very rare. In a meta-analysis, Cauchoix et al. (2018) identified only six publications reporting reliability for any measure of cognitive performance, and of these only two measured innovative problem-solving in wild subjects and both were in birds. Thus, there is a pressing need to examine the reliability of innovative problem-solving in other wild animals, especially in wild mammals. Furthermore, most studies only measure cognition at two time points and across two to four different tasks (Ashton, Ridley, Edwards, & Thornton, 2018; Cole, Cram, & Quinn, 2011; McCune, Jablonski, Lee, & Ha, 2019; Shaw, 2017) and there has been very little research examining how reliability might vary based on the number of measures or the length of the interval between measures, how reliability might vary across different environmental contexts, nor how reliability might vary among different age-sex classes [e.g. 13]. Our ignorance here is due, in part, to the numerous logistical challenges of experimentally measuring innovation repeated times in the same subjects— a problem that is particularly pronounced in wild subjects where locating and enticing individuals to perform cognitive tests even once can be difficult and tracking individuals for repeated tested may be nearly impossible in many species. However, for the intraspecific approach in wild animals to be useful for testing evolutionary hypotheses, researchers *must* provide evidence that measures of innovative-problem solving reflect stable traits and that estimates of reliability are robust against numerous sources of variation in testing environment and methodology.

The spotted hyena is a well-established as a model organism for testing hypotheses about the evolution of cognition (Holekamp et al., 2007) and innovativeness has previously been measured in both captive and wild hyenas (Benson-Amram & Holekamp, 2012; Benson-Amram et al., 2013; Johnson-Ulrich, Johnson-Ulrich, & Holekamp, 2018), but the problem-solving paradigms used to measure innovation were never previously tested for reliability except by Johnson-Ulrich et al.(Johnson-Ulrich, Benson-Amram, & Holekamp, 2019). Here, we measured reliability (R) by calculating intraclass correlation coefficients (ICC), which are commonly used in behavioral ecology to assess the reliability of behavioral traits within individuals (Nakagawa & Schielzeth, 2010). An ICC estimates the amount of variation in the response variable explained by random effects or grouping factors in mixed hierarchical models. Ultimately, we found a high level of reliability for innovative problem-solving in wild spotted hyenas and demonstrate how estimates of reliability vary across task demands, age-sex classes, socio-ecological contexts, the temporal interval between observations, and the total number of observations.

METHODS

Study site and subjects

We tested innovation in four neighboring spotted hyena clans within the Maasai Mara National Reserve, Kenya between June 2016 and November 2017. These clans ranged in size from 30 to 55 adult hyenas. Spotted hyena clans represent distinct social groups that are made up of multiple unrelated females, their offspring, and adult immigrant males. Clans are structured by strict linear dominance hierarchies, with an alpha female and her offspring at the top, followed by lower-ranking females and their offspring, with adult immigrant males ranking at the bottom. Births occur year-round and unrelated females raise their offspring together at a communal den. Female hyenas stay in their natal clan throughout their lives, whereas male hyena usually disperse to join new clans when they are 24-60 months old, after they reach sexual maturity (Engh, Esch, Smale, & Holekamp, 2000; Van Horn, McElhinny, & Holekamp, 2003).

All subjects were identified by their unique spot patterns and ear damage. Hyenas of all age classes and both sexes were included in the study. All subjects were sexed within the first few months of life based on the morphology of the genitalia (Smith & Holekamp, 2018). Age classes were based on life history stage (Holekamp & Dloniak, 2010). Cubs were defined as hyenas that were still dependent on the communal den; on average, Mara cubs become den-independent around 9-12 months of age (Holekamp & Dloniak, 2010). Subadults were hyenas who were den-independent but had not yet reached sexual maturity. Adults were hyenas that had reached sexual maturity. In females, sexual maturity was determined by the observation of mating, visual evidence of first parturition, or the female reaching three years of age, whichever came first (Holekamp, Smale, & Szykman, 1996). In males, sexual maturity was determined by dispersal status, males who were still present in their natal clan at testing were classified as subadults and immigrant males were classified as adults.

Multi-access box paradigm for measuring repeated innovation

We tested innovation in wild spotted hyenas using a multi-access box designed for use with mammalian carnivores (Johnson-Ulrich et al., 2018). The multi-access box (hereafter, ‘the MAB’) is a problem-solving paradigm, also known as an artificial or novel, extractive foraging task, where subjects must solve a novel problem to obtain a food reward. In contrast to traditional problem-solving tasks, MAB paradigms typically offer multiple solutions to the same puzzle box, each requiring its own unique behavior pattern. As a condensed battery of tasks, the MAB paradigm allows researchers to measure innovation, not just once, but multiple times across different doors (Auersperg et al., 2012).

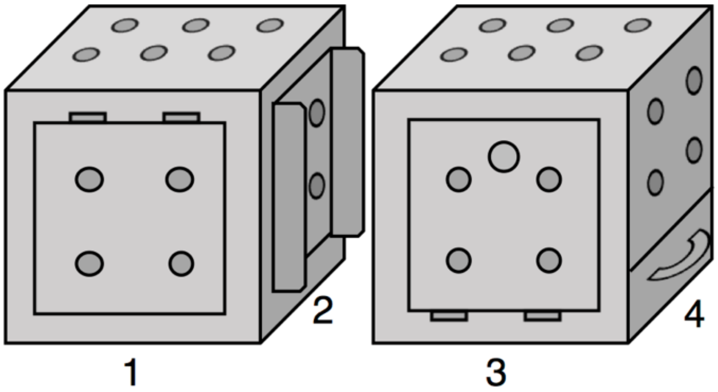

Figure 3.1. The MAB used in the current study. 1) the push flap solution; 2) the sliding door solution; 3) the pull flap solution; and 4) the drawer solution. Small filled grey circles indicate the approximate number and location of holes drilled through the wall of the MAB. Large gray circle on side 3 represents the location of the doorknob. Small rectangles represent the location of door hinges.

We chose to use a MAB paradigm because it allowed us to compare reliability across repeated trials within the same door to reliability across different doors. Reliable success with the same door across trials may be a result of individual learning rather than a result of a stable cognitive trait. However, if individuals reliably innovate by opening multiple unique doors to the MAB this would suggest that innovation is a stable cognitive trait. The MAB in the current study was a steel box with four unique doors, each requiring a different motor behavior, that could be used to access a common interior baited with a food reward (Figure 3.1). This MAB was previously used to test repeated innovation in captive hyenas; for more information about the design specifications see Johnson-Ulrich et al. (Johnson-Ulrich et al., 2018).

Test protocol

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

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

Test trials followed a similar protocol to familiarization trials; a trial began when a hyena came within 5 m of the MAB and ended upon successful opening of one of the solutions to the MAB, a “success”, or when the hyena moved more than 5m away from the MAB for more than 5 minutes. If the hyena contacted the MAB, but failed to open it, the trial was considered to be a “failure” for that hyena. Hyenas who approached the MAB, but did not make contact, were recorded as not participating in the trial and were not assigned success or failure. Average duration of test trials was 13.5 ± 15.7 minutes.

If a hyena had a “feed” familiarization trial or successful test trial, and if it had moved at least 5m away from the MAB, the MAB was immediately rebaited for successive testing. Hyenas were given successive trials as long as the testing conditions remained suitable, as described above, or until researchers ran out of bait. Successive trials were not administered following trials where every hyena that participated was unsuccessful because unsuccessful hyenas were those that had moved beyond 5m from the MAB for more than five minutes without opening the MAB and these hyenas were extremely unlikely to spontaneously re-approach the MAB for another trial. On average, hyenas received 1.53 ± 1.25 trials per testing session and completed testing across 6.31 ± 2.58 separate sessions (Figure 3.2). Most sessions were separated by a median of 1 day (mean ± SD = 24.18 ± 60.30 days, range = 0 – 321 days).

Test trials were divided into four different phases of testing. During Phase 1, the MAB was presented to hyenas with all four doors accessible. After a hyena had reached completion criterion for phase 1, defined by success with the same door in three out of four consecutive trials, it progressed to Phase 2. During Phase 2, the door used in Phase 1 was blocked by bolting it shut. The same criteria for progression applied to subsequent phases until a hyena had reached the criteria for progression with all four doors. Hyenas were given trials until they either reached criterion for all four doors or failed five consecutive trials during any phase of testing. Hyenas that participated in less than five trials, of which none were successful, were not included in our analysis. On average, hyenas 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 trials (N = 15), phase 3 in 4.08 ± 1.32 trials (N = 13) and phase 4 in 4.25 ± 1.96 trials (N = 12).

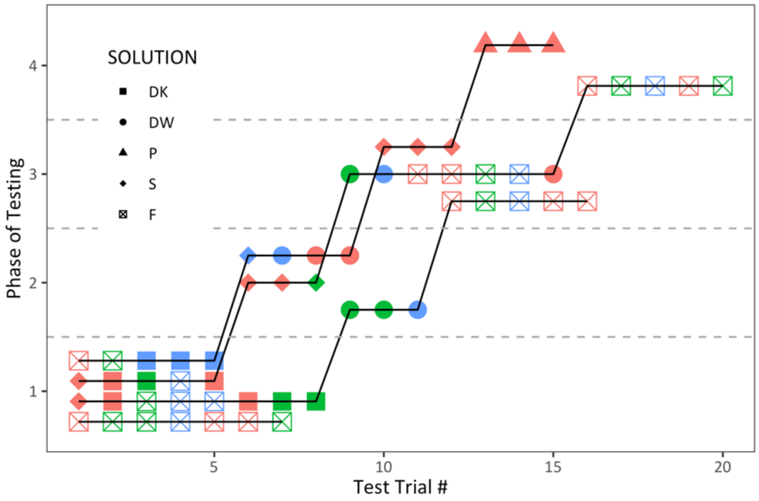

Figure 3.2. Visualization of test trial results for four hyenas across the four phases of testing. Each line represents a single hyena. Shapes indicate the door used to open the MAB (DK = door knob door; DW = drawer, S = sliding door, P = push door) or failure (F) to open any of the four doors on each trial. Adjacent shapes with the same color indicate trials that took place within the same testing session whereas adjacent shapes with different colors indicate trials that were separated in different testing sessions. Test trial # indicates the temporal order in which trials occurred; trials within the same session occurred immediately following previous trials, but the amount of time between trials in different sessions varied.

We aimed to give every hyena two familiarization trials prior to being given the option to participate in test trials. On average hyenas were given the opportunity to participate in 1.60 ± 1.54 (mean ± standard deviation) familiarization trials prior to their first phase one trial, but hyenas only fed from the MAB on and average of 0.94 ± 1.11 familiarization trials prior to their first phase one trial.

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

Statistical Analysis

All statistical analyses were performed using the statistical software R (R Core Team, 2019). Here, R values were calculated for subject ID in generalized linear mixed models (GLMMs). The rptR package also allowed us to estimate uncertainty around each point estimate for R via parametric bootstrapping (n = 1000), in which we estimated a standard error, a 95% confidence interval, and a *P*-value for each estimate of R. *P*-values were generated using likelihood ratio tests where model fit was compared to a null model with no grouping factor. Here, R values were calculated as a ratio of the variance explained by the subject ID over the residual variance (adjusted R values). Variance explained by fixed effects was not included in the denominator.

Before calculating R for innovation, we created a global model that included age class, sex, rank, and clan as fixed effects and subject ID as a random effect in order to identify factors that might influence innovation. We used the glmmTMB package to create all global models (Brooks, Mollie et al., 2017). To identify fixed effects of importance we used the ‘dredge’ function in the R package MuMIn (Bartoń, 2018). Dredge identified five top models with a delta AICc of less than four (Table 3.A.2). However, none of the fixed effects included in top models were significant, but the push door, was marginally more difficult than the other three doors (Binomial GLMM: $z = -1.74$, $P = 0.08$) and we therefore decided to include a fixed effect indicating the door used in our rptR models testing the reliability of innovation (Table 3.1: Models 1.1 and 1.2).

Next, before calculating R for problem-solving ability, we created a second global model that included age class, sex, rank, clan, trial number, and phase of testing as fixed effects, and subject ID as a random effect, to identify factors that might influence successful problem-solving. To identify fixed effects of importance we used the ‘dredge’ function in the R package MuMIn (Bartoń, 2018). Dredge identified six top models with a delta AICc of less than four (Table 3.A.3). Here, only trial number and phase number were included in all six top models, and only trial number and phase number had significant effects. Phase of testing had a significant negative effect on the likelihood of solving the MAB (Binomial GLMM: $z = -2.79$, $P < 0.01$) which suggests that later phases of testing, where solutions that hyenas use previously were blocked, were indeed more demanding for hyenas. After controlling for the effect of phase, overall test trial number had a significant positive effect on the likelihood of solving the MAB (Binomial GLMM: $z = 2.56$, $P = 0.01$), suggesting that hyenas were more likely to solve the MAB in later than earlier trials (Figure 3.3A).

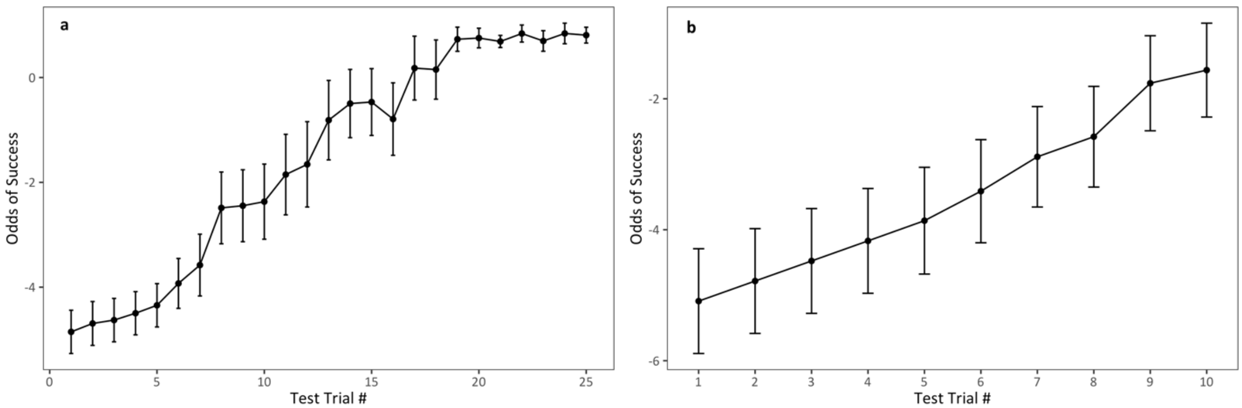

Figure 3.3. The predicted log odds of a successful trial with the MAB as a product of trial number from the binomial GLMM (Model 3). Error bars indicate standard error. **A)** Shows the log odds of successfully opening the MAB as a function of test trial number. **B)** shows the log odds of successfully opening the MAB as a function of trial number only for hyenas with at least 10 trials. A log odds of zero indicates equal odds of successful and unsuccessful trial, a negative log odds indicates a greater probability of an unsuccessful trial and a positive log odds indicates a greater probability of a successful trial.

The positive effect of trial number could indicate that hyenas were learning how to improve their performance across trials, but this effect might also be biased by unsuccessful subjects reaching the criterion to end testing (five unsuccessful trials in a row) and dropping out of the subject pool. To test this possibility we created another model where we restricted our dataset to the first ten test trials only for hyenas that had at least 10 trials, and found that test trial still had a significant positive effect on the likelihood of solving the MAB (Binomial GLMM: $z = 2.90$, $P < 0.01$, n = 428 trials, N = 25 hyenas; Figure 3.3B). Therefore, we included both test trial number and phase of testing where appropriate in all subsequent rptR models that assessed the reliability of problem-solving success across trials.

RESULTS

Seventy-two hyenas participated in 694 test trials with a multi-access box (MAB). Overall, our sample included an adequate representation of each age and sex class with 17 adult females, seven adult males, 13 subadult females, 17 subadult males, 10 female cubs, and nine male cubs. Out of these 72 hyenas, 23 opened the MAB at least once (mean = 2.74 doors, SD = 1.39) and 11 opened each of the four doors to the MAB at least once across their trials. Across all hyenas' trials, a mean of 0.28 ± 0.45 (mean ± SD) were successful. Overall success rates on trials were lowest during phase one of testing (0.15 ± 0.36; N = 72), but, for hyenas that progressed, success rates increased on phase two (0.93 ± 0.26; N = 15), phase three (0.79 ± 0.41; N = 13), and phase four of testing (0.45 ± 0.50; N = 12).

Reliability of innovation

We first investigated the reliability of innovation by examining innovation across the four different MAB solutions (door knob, slot, push, drawer). Because innovation is defined as solving a novel problem, hyenas only demonstrated innovation when they solved each door of the MAB for the very first time. Therefore, we measured innovation as a binary variable indicating whether or not a hyena had solved each of the four particular doors to the MAB at least once (Table 3.1: Model 1.1 and Model 1.2) regardless of phase of testing. Each hyena received four dichotomous scores for each of the four unique MAB doors, with a score of one indicating that they solved a door at least once and a score of zero indicating that they never solved that particular door despite contacting the MAB on multiple trials. We included door ID as a fixed effect in this model (See Methods: Statistical Analysis). Reliability was determined by calculating ICC in the R package rptR (Stoffel, Nakagawa, & Schielzeth, 2017).

We found a high level of reliability for innovation across all subjects (Likelihood ratio test: $R = 0.97$, $P < 0.001$; Table 3.1: Model 1.1). However, two-thirds of all hyenas never innovated with any of the four MAB doors (Figure 3.4). Because high reliability here could be driven by this zero-inflation, we next calculated reliability only within subjects who had solved the MAB at least once. Among these 23 hyenas, we found that innovation was moderately reliable (Likelihood ratio test: R = 0.48; *P* = 0.002; Table 3.1: Model 1.2).

Table 3.1. Results from rptR models on the reliability of innovation.

	Model Description	**R**	**SE**	**95% CI**	***P*-adjusted**	**N**	**n**
Model 1.1	Innovation – all subjects	0.97	0	0.97,0.97	<0.001*	72	288
Model 1.2	Innovation – just innovators	0.48	0	0.48,0.48	0.002*	23	92
Model 2.1	Phase 1 door knob	0.54	0.21	0.08,0.94	0.001*	23	81
Model 2.2	Phase 1 sliding door	0.53	0.30	0.02,0.97	0.021*	23	81
Model 2.3	Phase 1 drawer	0.36	0.30	0,0.96	0.236	23	81
Model 2.4	Phase 1 push door	0.69	0.26	0.14,0.99	0.009*	23	81

Note: *P*-values adjusted using the Bonferroni method for N = 33. 'N' is the number of subjects and 'n' is the number of trials (Model 1.1 and 1.2) or the number of successful phase one trials (Model 2.1 to 2.4).
*Indicates statistical significance at α = 0.05

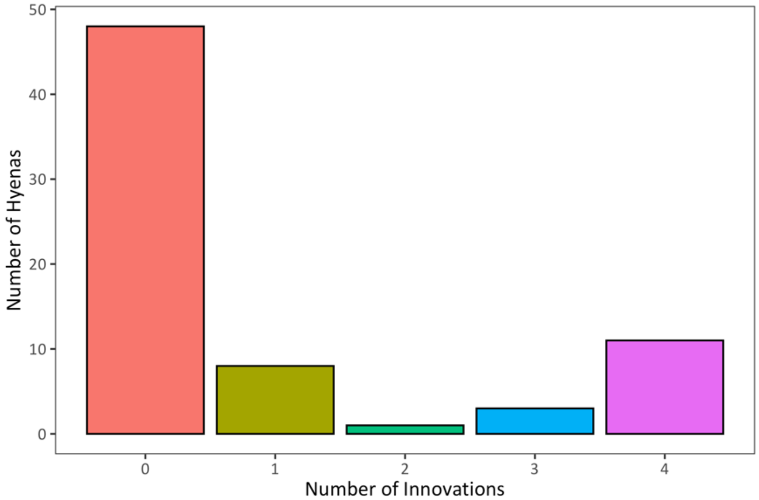

Figure 3.4. Count of the number of hyenas obtaining each innovation score for the number of times they were able to innovate with the MAB (N = 71 hyenas).

Next, we evaluated the reliability of the variable reflecting hyenas' use of the same door across trials within phase one of testing when all four doors were available to them. If hyenas consistently opened the MAB with the same door during phase one, despite the availability of other doors, this might suggest they were learning the affordances of that door.

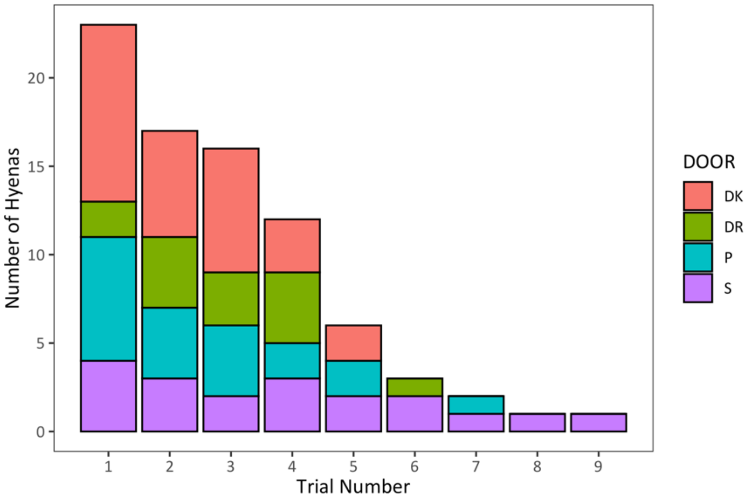

Figure 3.5. The number of hyenas using each solution across their successful trials within phase one of testing. Twenty-three hyenas were successful on at least one trial, but within these twenty-three hyenas the number of successful trials they had within phase one varied. Abbreviations: door knob (DK), drawer (DW), push door (P), sliding door (S).

We used a dataset that included hyenas' scores for each trial but only included successful, phase one trials. Because door type was a multinomial response variable, we created four binomial models to determine the reliability of each variable separately. Each response variable was binary and indicated whether a hyena did or did not use a specific door when they solved the MAB. We found that hyenas tended to be moderately consistent at using three of the doors in Phase 1 (Likelihood ratio test: R = 0.53 – 0.69; *P* = 0.001 – 0.021; Table 3.1: Models 2.1-2.4), which suggests that if they solved the MAB using a particular door during phase one, they tended to keep using that same door throughout phase one (Figure 3.5). The exception to this was the drawer solution, which had nonsignificant reliability (Likelihood ratio test: R = 0.36, *P* = 0.236; Table 3.1: Model 2.3).

Reliability of problem-solving

In addition to evaluating the reliability of variables reflecting innovative ability, we were also interested in investigating problem-solving success by hyenas on each individual trial. We gave each hyena multiple trials with the MAB in order to give subjects the opportunity to solve different doors of the MAB. This dataset gave us the opportunity to look at the reliability of problem-solving across different age-sex classes, different contexts, timespans, and testing protocols. Because hyenas tended to consistently use the same door across trials within Phase 1, with the exception of the drawer solution (Table 3.1), in all following models our response variable was a binary variable indicating whether a hyena opened or failed to open a door of the MAB, irrespective of which specific door they solved. We included trial number and phase of testing as fixed factors in these models to control for the effect of experience and varying task demands across phases (See Methods: Statistical Analysis).

Reliability across all trials

In Model 3, we first assessed reliability of problem-solving within subjects on all trials across time and across all contexts to estimate overall reliability. We found that performance across all trials was significantly highly reliable (Likelihood ratio test: R = 0.91, $P < 0.001$; Figure 3.6; Table 3.A.4: Model 3).

Reliability within different age and sex classes

Next, we were interested in whether reliability differed among individuals in different age sex classes. For example, some literature suggests that males exhibit more reliable behavior than females (Bell et al., 2009) and it seems reasonable to expect that juveniles, who are still developing, might exhibit behavior that is less reliable than that of adults. To compare reliability within different age and sex classes we partitioned our dataset into six different categories: adult female, subadult female, adult male, subadult male, and cubs. Data for male and female cubs was pooled because our sample size was too small to accurately estimate R for each sex independently, and because most behavioral sex differences in this species emerge after cubs become independent of dens, e.g. (Boydston, Kapheim, Van Horn, Smale, & Holekamp, 2005). Likewise, the data for adult males was insufficient for using bootstrap analysis to estimate standard error or confidence intervals. We found that all age-sex classes had generally high reliability (Likelihood ratio test: R = 0.78 – 0.94, $P < 0.001$; Figure 3.6; Table 3.A.4: Models 4.1 – 4.5;).

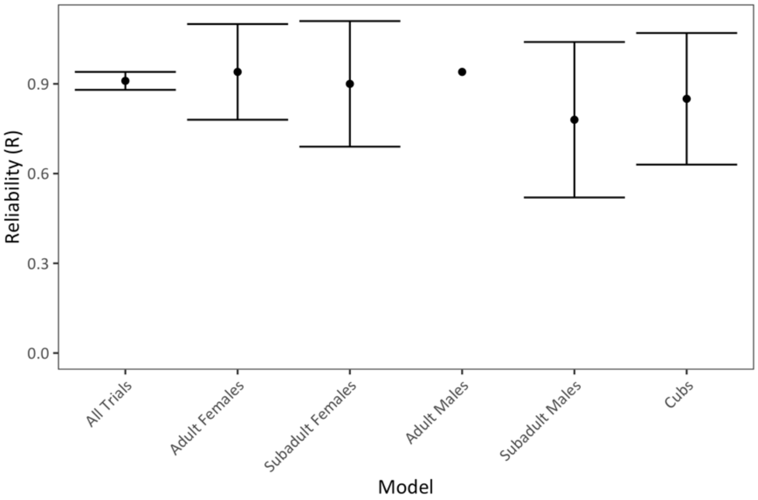

Figure 3.6. Reliability estimates for problem-solving success within different age-sex classes. Error bars show standard error calculated using parametric bootstrapping (N = 1000). Sample sizes were too small to estimate standard error for adult males.

Cross-contextual reliability: Phase of testing

Because the task demands change across different phases of testing, we were next interested in the reliability of hyenas' performance across these different phases. In phase one, all four possible doors to the MAB are available to hyenas, but in subsequent phases all doors that a hyena used previously are blocked, forcing the hyena to use a different solution and inhibit persevering at blocked doors. Out of all 72 hyenas, 12 hyenas progressed to phase four of testing and 11 hyenas were ultimately able to innovate with four different doors, only seven of these 12 hyenas actually reached criterion for completion of phase four (see Methods: Test protocol). First, we looked at reliability within each phase of testing by including a nested random effect of phase in our model. We found that reliability was high within phases of testing (Likelihood ratio test: R = 0.91, *P* < 0.001; Table 3.2: Model 5.1).

Table 3.2. Reliability within and across different contexts.

	Model Description	**R**	**SE**	**95% CI**	***P*-adjusted**
Model 5.1	Within Phases	0.91	0.03	0.85,0.98	<0.001*
Model 5.2	Across Phases (pass/fail)	0.98	0	0.98,0.98	<0.001*
Model 6.1	Within Seasons	0.91	0.04	0.85,0.98	<0.001*
Model 6.2	Across seasons (proportion)	0.77	0.28	0.16,0.96	0.018*
Model 7.1	Within locations	0.91	0.04	0.85,0.98	<0.001*
Model 7.2	Across locations (proportion)	0.80	0.24	0.18,0.97	0.066
Model 8.1	Within group sizes	0.92	0.03	0.87,0.98	<0.001*
Model 8.2	Across group sizes (proportion)	0.81	0.15	0.26,0.97	<0.001*

Note: *P*-values adjusted using the Bonferroni method for N = 33. All models included all 71 subjects and 685 trials.
*Indicates statistical significance at $\alpha = 0.05$.

Next, we investigated whether a hyena's performance in one phase of testing was consistent with its performance in other phases of testing. We summarized performance in each phase with a binary variable that indicated whether or not a hyena completed each phase by reaching the criterion to progress. We found that whether or not a hyena passed each phase of testing was highly consistent across phases (Likelihood ratio test: R = 0.98, *P* < 0.001; Table 3.2: Model 5.2). This suggests that hyenas were generally consistent in their ability to reach the criterion for completing a phase despite changing task-demands.

Cross-contextual reliability: Environmental factors

Because environmental factors can influence problem-solving success, we examined the reliability of problem-solving across two different environmental conditions, the season of testing and the location of testing. As above, these models looked at reliability both within contexts and between contexts as a count of successful trials relative to the total number of trials. We considered two different seasons of testing: during either the presence or absence of the annual wildebeest migration. The annual wildebeest migration in the MMNR represents a massive influx of prey for spotted hyaenas (Holekamp, Smale, Berg, & Cooper, 1997) and this could potentially affect motivation to obtain a food reward from the MAB. The arrival and departure of the wildebeest migration in the Maasai Mara National Reserve varies from year to year, but the presence of the migration there generally spans the months of July through October (Sinclair & Norton-Griffiths, 1984). We counted the migration as present if there was at least one herd of wildebeest inside the territories of our study clans during each week of testing. First, we computed reliability within each season by including a nested random effect of season in our model. We found that reliability within seasons was high (Likelihood ratio test: $R = 0.91$, $P < 0.001$; Table 3.2: Model 6.1). Next, in order to compare performance across seasons we analyzed performance within each season as the proportion of successful trials within each season (Table 3.2: Model 6.2). We found that problem-solving performance was moderately reliable across seasons (Likelihood ratio test: $R = 0.77$, $P = 0.018$; Table 3.2: Model 6.2).

Regarding test location, we considered two locations that are socially important for spotted hyenas: the communal den and the carcasses of prey animals. Both location types typically attract larger numbers of hyenas than do other areas of a clan's territory, but the communal den is a static location representing a central hub for hyena socializing, whereas the locations of carcasses are variable and feeding competition is high at them. We coded location with three different values: den, carcass, and other. We recorded the location as ‘den’ if a communal den was within 100 m of the MAB. ‘Carcass’ was recorded as the location if hyenas were feeding within 100 m of the MAB on visible animal remains, other than those provided by researchers for testing. If neither the den nor a carcass was within 100m of the MAB, the session’s location was recorded as ‘other.’ First, we computed reliability within all locations by including a nested random effect of location type in our model. We found that reliability was generally high within each location (Likelihood ratio test: R = 0.91, $P < 0.001$; Table 3.2: Model 7.1). Next, for subjects who had trials across multiple locations we compared performance within one location to their performance in other locations by analyzing performance within a location as the proportion of successful trials within each location (Table 3.2: Model 7.2). We found that problem-solving performance across different locations was nonsignificant for reliability (Likelihood ratio test: R = 0.80, $P = 0.066$; Table 3.2: Model 7.2) which suggests that hyenas do not interact with the MAB consistently across different locations.

Cross-contextual reliability: Social factors

Because social factors can also affect foraging behaviors and create substantial distractions for test subjects, we also examined how reliable performance was within and across the size of testing group during each trial. Although we attempted to only test groups of hyenas that numbered five or less, we were not able to control for additional hyenas leaving or arriving during trials. On average, the number of hyenas participating in a trial was 3.89 ± 3.71 (median = 3, range = 1 – 29). In order to compare reliability across different group sizes we binned group size into small (1 – 2), medium (3 – 5), and large (6 – 29) groups. We found that performance was highly reliable within each group size (Likelihood ratio test: R = 0.92, $P < 0.001$; Table 3.2:

Model 8.1) and across each group size (Likelihood ratio test: R = 0.81, *P* < 0.001; Table 3.2: Model 8.2).

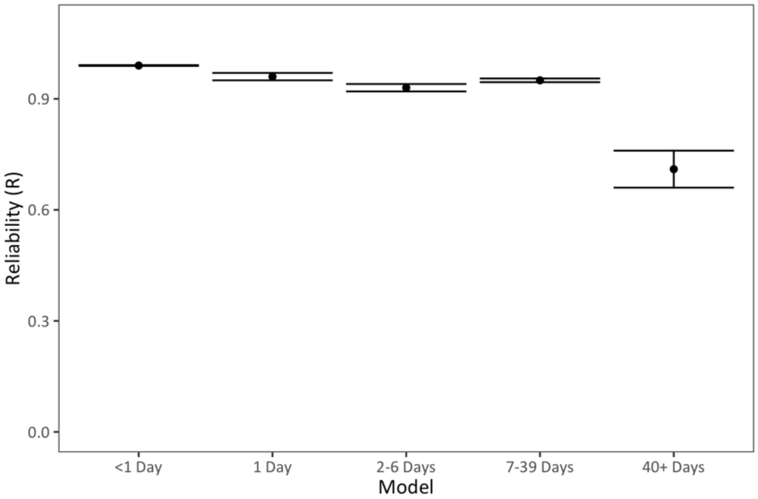

Figure 3.7. Reliability estimates of problem-solving success across different timespans between pairs of trials. Error bars show standard error calculated using parametric bootstrapping (N = 1000).

Temporal reliability

Although most test trials were conducted only 1 day apart for subjects, the average number of days between testing sessions was 24.18 ± 60.27 days (median = 1, range = 0 – 321 days). We were interested in whether reliability between any given trial and the trial that followed it was affected by the amount of time between trials. To do this, we created a dataset where we paired each subject's trial with the trial that followed it and calculated the number of days elapsing between the two trials. We next partitioned this dataset into trials that occurred less than one day apart, exactly one day apart, two to six days apart, seven to 39 days apart, and 40 or more days apart. We then calculated reliability between pairs of trials for each of these datasets (Table 3.A.5; Figure 3.7). We found that reliability remained high even for trials separated by 7-39 days (Likelihood ratio test: R = 0.95, *P* < 0.001; Table 3.A.5: Model 9.4), but reliability became non-significant when trials were separated by 40 or more days (Likelihood ratio test: R = 0.72, $P = 0.219$; Table 3.A.5: Model 9.5).

Reliability across different numbers of trials

Finally, we were interested in how the varying number of trials collected per hyena might affect estimates of reliability. On average, hyenas received 9.64 ± 5.61 trials (median = 7 trials, range = 2 – 26 trials). Collecting a greater number of trials per hyena could, in theory, increase the accuracy of estimates about their innovative problem-solving ability and therefore increase reliability. However, because we were testing hyenas in the wild, larger number of trials were more likely to take place across different testing sessions, different timespans, or different contexts. Although we generally found that temporal and cross-contextual reliability was high, we were interested in how our estimates might have changed if we'd only sampled hyenas a set number of times. To do so, we calculated reliability for hyenas in nine models where we included only their first two to ten trials. We found that reliability was extremely high within a hyena's first two and first three trials (Likelihood ratio test: R = 0.88 – 0.93, $P < 0.001$; Figure 3.8; Table 3.A.6: Models 10.1 – 10.2) and that reliability was lowest when we only included each hyena's first four trials (Likelihood ratio test: R = 0.41, $P < 0.001$; Figure 3.8; Table 3.A.6: Model 10.3). The inclusion of subsequent trials after the fourth trial appeared to generally increase estimates of reliability, however, the confidence interval for these estimates overlapped that of the estimate of reliability for the first four trials, making it difficult to draw the conclusions about the benefits of added trials. Therefore, we ran one additional model where we pooled the data from Models 10.4 through 10.9 in order to estimate the reliability of problem-solving across the first ten trials for hyenas who received at least five trials. This dataset included up to ten trials per hyena and only included less than ten trials for hyenas that participated in five to nine trials. With this data pooled, we found a high level of reliability for hyenas' first 5-10 trials (Likelihood ratio test: R = 0.89, *P* < 0.001; Table 3.A.6: Model 10.10).

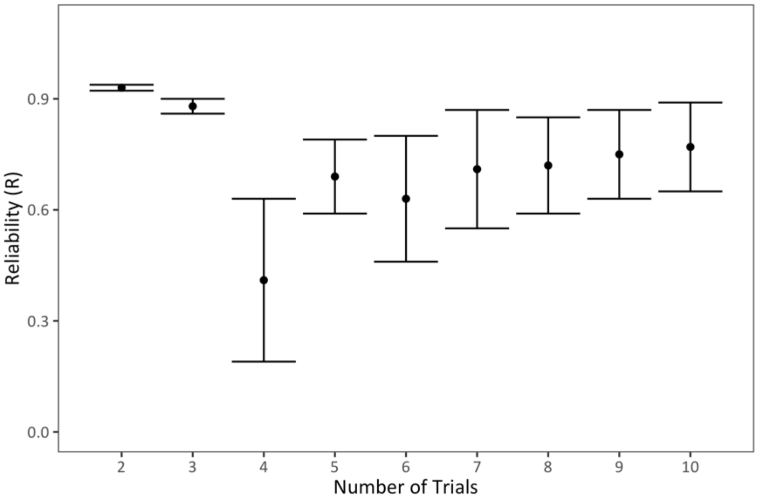

Figure 3.8. Reliability estimates for problem-solving success across different numbers of trials. Error bars show standard error calculated using parametric bootstrapping (N = 1000).

DISCUSSION

Overall, our results suggest that innovativeness and problem-solving ability are stable, general traits in wild spotted hyenas. Compared to the average reliability of behaviors in wild animals (Bell et al., 2009), both of these traits exhibited remarkably high levels of reliability. These results generally agree with other studies of innovative problem-solving in captive and wild animals that have also found fairly high levels of reliability (Ashton, Ridley, et al., 2018; Borrego & Dowling, 2016; Daniels, Fanelli, Gilbert, & Benson-Amram, 2019), but we also present the evidence demonstrating that innovative problem-solving is generally reliable across age-sex class, task demands, socio-ecological contexts, the interval between observations, and the number of observations, with a few important exceptions.

We found high levels of reliability for innovativeness across the four different MAB doors. These doors represent four different motor tasks designed to measure innovation. To ensure that this high reliability was not solely explained by some hyenas consistently failing to innovate, we also computed the reliability of innovation among only hyenas that had innovated at least once. Here, we found a moderate level of reliability, hyenas who innovated at least once were moderately likely to demonstrate repeated innovations (Table 3.1). This result is similar to studies in wild and captive birds that have generally found consistent performance among problem-solving tasks requiring different motor actions (Cole et al., 2011; Ducatez et al., 2015; Lermite, Peneaux, & Griffin, 2017; McCune et al., 2019). We also found that hyenas were moderately consistent at using the same door during Phase 1 of testing (Table 3.1). The only exception to this finding was with the drawer door, for which hyenas' performance was not reliable. We suspect performance with the drawer was not reliable because the motor behavior most hyenas used to open it, biting and pulling, was similar to the motor pattern used to open the door knob (also biting and pulling) with one key difference: to open the drawer a hyena was required to position their head extremely close to the ground and with the door knob a hyena only had to lower their head slightly (adults) or not all (subadults and cubs). Thus, it is possible that hyenas who used the drawer might have retained the motor behavior while switching to using the more easily accessible location. In contrast, performance with the push door had the highest reliability out of all four solutions. This solution also differed the most strongly from the other three solutions in terms of the motor behaviors required to solve it; the push door could only be opened by depressing the door inwards while the other three solutions could only be opened outwards. This difference may have contributed both to its reliability and its difficulty; the push door had the lowest success rate out of all four doors (GLMM: $z = -1.70$, $P = 0.09$).

In addition to evaluating the reliability of innovativeness in spotted hyenas, we were also curious how consistent hyenas were at solving the MAB on each trial, regardless of the door used. We found remarkably high reliability for problem-solving performance across all trials. We also found high reliability among each age-sex class we examined. These results differ from results for behavior across animals more generally, a meta-review of the reliability of animal behavior found that males tend to be more reliable than females (Bell et al., 2009).

Reliability for problem-solving was also high within and across different contexts including the phase of testing, the season of testing, location of testing, and social context. Innovative problem-solving was highly reliable across the four different phases of testing despite the changing task-demands across phases. We were also surprised to find high reliability for problem-solving between seasons because previous research suggests the presence or absence of the wildebeest migration can have large effects on the foraging behavior and space-use of spotted hyenas (Hofer & East, 1993). The nonsignificant reliability for problem-solving performance across different locations was unsurprising because it is highly likely that motivations and distractions differ between these different locations. For example, assuming a carcass is a more valuable food resource than the small amounts of bait provided inside the MAB, hyenas might be less motivated to solve the MAB when a carcass is present. In addition, the higher degree of socializing at communal dens might facilitate participation in trials with the MAB at communal dens. Previous research found that social facilitation plays a strong role in the way hyenas interact with a puzzle box (Benson-Amram et al., 2014). However, we also found high reliability across different social contexts; hyenas show consistent performance across small, medium, and large group sizes.

In general, estimates of reliability are higher for behavioral observations that are closer together in time (Bell et al., 2009). Here, we found remarkably high reliability of problem-solving performance even for pairs of trials separated by over a month. Only when trials were separated by 40 days or more did we find no reliability between them. This is an important finding because it can be extremely difficult to consistently locate subjects for repeated testing, especially in animals like spotted hyenas that live in fission-fusion societies occupying large territories.

In a meta-review of reliability in earlier animal behavior research, reliability estimates were generally not affected by the number of observations per individual (Bell et al., 2009). Similarly, we found that problem-solving performance was significantly reliable regardless of the total number of observations taken per individual hyena. However, although reliability was very high when we only analyzed each hyena’s first two or three trials, reliability decreased when we included each hyena’s first four trials before increasing again with more than four trials. This finding was initially surprising, and it may partially result from our testing protocol. On average hyenas received 1.58 ± 1.33 trials per testing session, so their fourth trial is slightly more likely to take place during a different testing session and thus may be more likely to occur in a different context or be separated from earlier trials by varying amounts of time. In addition, three trials was the minimum number required before progressing to phase two and some fourth trials may represent subjects’ first trial in phase two. Therefore, the same factors affecting lower levels of reliability across different contexts and timespans might also explain lower reliability when more problem-solving trials were included in our analysis. This result also potentially implies that high reliabilities obtained when researchers test animals only two or three times may be artificially inflated as a result of similar testing conditions, and that a small number of trials may not provide adequate opportunity for subjects to succeed at cognitive tasks, especially if the trials are conducted under similar conditions and close in time. This result is important because an important weakness of the design of many cognitive tests is that a subject only needs to succeed at a cognitive task once to demonstrate an ability, but the opposite is not true; failure at a cognitive task is usually not diagnostic of a lack of an ability (Byrne, 1995b; Shettleworth, 1998). Because reliability was lowest when we only analyzed each subject's first four trials, and because the fourth trial was the highest trial in which any of our subjects ever innovated for the first time (mean ± SD = 1.94 ± 1.25, range = 1 – 4), we used a conservative criterion of at least five consecutive failures in order to classify a hyena as non-innovative. Indeed, the high reliability estimates we obtained for innovation were only possible through repeated testing and nearly all 'innovative' hyenas failed at least one trial (Likelihood ratio test: $R = 0.48$, $P < 0.001$; Table 3.1, Model 1.2), Here, we suggest that, if researchers are testing wild subjects, they should aim to collect trials over a diverse array of testing conditions and identify an appropriate minimum number of trials per subject.

Our study offers an important demonstration of the reliability of innovative problem-solving in a wild mammal. However, high reliability does not necessarily correlate with high validity. Previous research has heavily debated the conceptual validity of problem-solving paradigms for measuring innovation (Griffin & Guez, 2014; Reader et al., 2016; Rowe & Healy, 2014; van Horik & Madden, 2016). Although this debate is not entirely settled, researchers have found that the behaviors underlying spontaneous innovation in the wild are very similar to the behaviors that underlie experimentally measured innovation using problem-solving paradigms (Griffin & Guez, 2014), which strongly suggests that problem-solving paradigms are valid measures of innovation. However, it is also important to consider the ecological validity of a paradigm and tasks should be designed with a species' underlying capabilities in mind. We designed a multi-access box that required spotted hyenas to use behaviors that are part of their natural foraging repertoire to solve a novel problem for a food reward. This kind of puzzle box is sometimes called a novel extractive foraging puzzle because it requires subjects to extract food from a container. Spotted hyenas are dietary generalists and mammalian bones, which represent an important part of their diets, require a moderate degree of extractive foraging to access the marrow within. Therefore, it is not surprising that spotted hyenas were able to innovate with this kind of problem-solving paradigm. However, for animals that never use extractive foraging in the wild, problem-solving paradigms like the one used in the current study might not be ecologically valid tools for assessing innovation.

In conclusion, it appears that, even with the many challenges posed by testing animals in the wild, we were still able to reliably measure innovativeness and problem-solving ability. However, the less reliable performance across different contexts reminds us that it is still important to test animals more than once or twice and preferably in different contexts, if only to provide adequate opportunities for animals to demonstrate, or fail to demonstrate, their abilities. Overall, our results complement the literature on the validity of innovation, and we conclude that problem-solving paradigms are reliable tools for measuring variation in innovative abilities.

APPENDIX

Table 3.A.1. Binomial GLMM model results on the factors affecting participation in MAB trials. Dependent variable was a binomial factor indicating whether a hyena participated in a trial or not.

Factor	**β**	**SE**	***z***	***P***
Park – Triangle/Reserve	0.23	0.24	0.94	0.35
Sex	-0.32	0.24	-1.33	0.18
Age class – cub	1.28	0.27	4.74	<0.001*
Age class – subadult	1.52	0.27	5.59	<0.001*
Rank	-0.29	0.12	-2.54	0.011*
# prior feeds	0.54	0.08	7.12	<0.001*
# prior solves	0.15	0.04	3.89	<0.001*
Trial #	-0.05	0.01	-5.41	<0.001*
Familiarization vs test trial	0.13	0.12	1.12	0.27
Body condition – fat	0.14	0.21	0.67	0.50
Body condition – gaunt	-0.21	0.62	-0.33	0.74
Body condition – bloody	-0.68	1.12	-0.61	0.54
Bait – milk powder	0.08	0.19	0.43	0.67
Bait – bone	0.05	0.15	0.31	0.76
Bait – muscle tissue	0.03	0.19	0.17	0.86
Bait – offal	0.06	0.21	0.27	0.79
Bait – spoiled	0.18	0.18	1.02	0.31
Location – carcass	0.55	0.37	1.48	0.14
Location – den	0.10	0.17	0.56	0.58
# hyenas present within 20m of the MAB	-0.17	0.02	-9.81	<0.001*
# hyenas present contacting the MAB	0.24	0.03	9.19	<0.001*
Migration present/absent	-0.06	0.16	-0.36	0.72

*Indicates statistical significance at $\alpha = 0.05$.

Table 3.A.2. Top models (Δ AICc < 4) from the dredge output for the global model on innovation (n = 281 doors, N = 71 hyenas).

Age Class	Clan	Rank	Sex	Door	AICc	Δ AICc
NA	NA	NA	NA	NA	167.55	0
NA	NA	NA	0.55	NA	169.47	1.92
NA	NA	0.09	NA	NA	169.58	2.03
NA	NA	NA	NA	-1.71 – -0.49	170.10	2.56
NA	NA	0.05	0.54	NA	171.52	3.98

Note: Each row contains the output for a single model. Values in cells show the estimated effect. A range of values is reported for door because it was categorical factors with four levels.
*Indicates statistical significance at α = 0.05.

Table 3.A.3. Top models (Δ AICc < 4) from the dredge output for the global model on problem-solving (n = 687 trials, N = 71 hyenas).

Age Class	Clan	Rank	Sex	Test Trial #	Phase	AICc	Δ AICc
NA	NA	NA	NA	0.15*	-0.90*	489.99	0
NA	NA	NA	0.86	0.15*	-0.90*	491.61	1.62
NA	NA	0.16	NA	0.15*	-0.90*	491.96	1.97
NA	NA	0.09	0.82	0.15*	-0.90*	493.53	3.64
NA	-1.76 – 1.11	NA	NA	0.14*	-0.86*	493.71	3.72
-0.39 – 0.23	NA	NA	NA	0.15*	-0.89*	493.90	3.91

Note: Each row contains the output for a single model. Values in cells show the estimated effect. A range of values is reported for age class and clan because they were categorical factors with three levels and four levels respectively.
*Indicates statistical significance at α = 0.05.

Table 3.A.4. Reliability across all trials and within each age-sex class.

	Model Description	R	SE	95% CI	*P*-adjusted	N	n
Model 3	All Trials	0.91	0.03	0.85, 0.98	<0.001*	72	694
Model 4.1	Adult Females	0.94	0.16	0.37,0.99	<0.001*	17	187
Model 4.2	Subadult Females	0.90	0.21	0.12,0.99	<0.001*	13	122
Model 4.3	Adult Males	0.94	NA	NA	<0.001*	7	62
Model 4.4	Subadult Males	0.78	0.26	0.04,0.98	<0.001*	17	162
Model 4.5	Cubs	0.85	0.22	0.06,0.99	<0.001*	19	161

Note: *P*-values adjusted using the Bonferroni method for N = 33. Sample size for adult males was too small to estimate standard error and confidence intervals for R. ‘N’ is the number of subjects and ‘n’ is the number of trials.
*Indicates statistical significance at α = 0.05.

Table 3.A.5. Temporal reliability results for pairs of trials.

	Model Description	**R**	**SE**	**95% CI**	***P*-adjusted**	**N**	**n**
Model 9.1	<1 day	0.99	0.002	0.99,0.99	<0.001*	56	256
Model 9.2	1 day	0.96	0.008	0.97,0.99	<0.001*	39	62
Model 9.3	2-6 days	0.93	0.007	0.97,0.99	<0.001*	57	120
Model 9.4	7-39 days	0.95	0.005	0.98,0.99	<0.001*	51	94
Model 9.5	>40 days	0.72	0.06	0.83,0.99	0.219	55	90

Note: *P*-values adjusted using the Bonferroni method for N = 27. 'N' is the number of subjects and 'n' is the number of trial pairs.

*Indicates statistical significance at α = 0.05.

Table 3.A.6. Reliability across different numbers of trials.

	Model Description	**R**	**SE**	**95% CI**	***P*-adjusted**	**N**	**n**
Model 10.1	First 2 Trials	0.93	0.008	0.97,0.99	<0.001*	72	144
Model 10.2	First 3 Trials	0.88	0.020	0.92,1	<0.001*	71	213
Model 10.3	First 4 Trials	0.41	0.22	0.18,0.89	<0.001*	71	285
Model 10.4	First 5 Trials	0.69	0.10	0.54,0.96	<0.001*	70	350
Model 10.5	First 6 Trials	0.63	0.17	0.35,0.94	<0.001*	58	347
Model 10.6	First 7 Trials	0.71	0.16	0.42,0.96	<0.001*	43	300
Model 10.7	First 8 Trials	0.72	0.13	0.42,0.94	<0.001*	32	255
Model 10.8	First 9 Trials	0.75	0.12	0.48,0.95	<0.001*	30	269
Model 10.9	First 10 Trials	0.77	0.12	0.46,0.95	<0.001*	25	248
Model 10.10	First 5-10 Trials	0.89	0.04	0.85,0.98	<0.001*	70	536

Note: *P*-values adjusted using the Bonferroni method for N = 33. 'N' is the number of subjects and 'n' is the number of trials.

*Indicates statistical significance at α = 0.05.

CHAPTER FOUR

GROUP SIZE AND SOCIAL RANK PREDICT INHIBITORY CONTROL IN SPOTTED HYAENAS

Johnson-Ulrich, L., & Holekamp, K. E. (2020). Group size and social rank predict inhibitory control in spotted hyaenas. *Animal Behaviour*, *160*, 157–168.
https://doi.org/10.1016/j.anbehav.2019.11.020

ABSTRACT

Inhibitory control is the ability to resist performing a prepotent, but ultimately incorrect, behaviour in situations that demand restraint. Inhibitory control is linked to brain size and intelligence in humans and animals, but it is unclear just how it evolves. Inhibitory control is thought to be particularly important in complex social environments where demands can shift frequently based on the social context and the identities or behaviours of other individuals in a group. Indeed, the social intelligence hypothesis suggests that the demands of living in complex social groups led to the evolution of sophisticated cognition. Here, we tested inhibitory control in wild spotted hyaenas, *Crocuta crocuta*, whose large social groups are structured by linear dominance hierarchies. We tested inhibitory control using the cylinder test, which requires subjects to inhibit going straight for a food reward. In support of our predictions, hyaenas living in larger groups had greater inhibitory control. In particular, the size of the cohort in which young hyaenas grew up, rather than the size of adult groups, had the strongest effect. In addition, the effect of group size was significantly stronger for low-ranking hyaenas, which must frequently inhibit both feeding and aggression in the presence of higher-ranking hyaenas. Contrary to our predictions, adult male hyaenas, which always occupy very low rank positions as adults, did not have better inhibitory control than adult females. This suggests that inhibition is

not a canalized trait, but instead may be a flexible one such that its development is influenced by early life social environments.

INTRODUCTION

Inhibitory control is the ability to resist performing a prepotent, but ultimately incorrect, behaviour in situations that demand restraint (Aron, 2007; Bari & Robbins, 2013). This cognitive skill is generally thought to be crucial to complex cognition, problem solving and behavioural flexibility (Diamond, 2013). It has been broadly studied in humans and other animals, and is strongly related to intelligence and measures of life success in humans (Diamond, 2013; Mischel et al., 1989) and to brain size in nonhuman animals (Horschler et al., 2019; Kabadayi, Taylor, von Bayern, & Osvath, 2016; MacLean et al., 2014). However, we know very little about the circumstances favouring the evolution of inhibitory control. One of the most popular hypotheses for the evolution of complex cognition is the social intelligence hypothesis (SIH), which suggests that the demands of living in complex social groups have led to the evolution of large brains and greater intelligence (Byrne & Whiten, 1988; Dunbar, 1998; Humphrey, 1976). Inhibition is thought to be particularly important in complex social environments where de- mands can shift frequently based on the social context and the identities or behaviours of other group members (Amici, Aureli, & Call, 2008). In social environments like these, restraining impulsive behaviour is necessary to maintain a stable or cohesive social group (Amici, Call, Watzek, Brosnan, & Aureli, 2018; Bjorklund & Harnishfeger, 1995; Byrne, 1995a; Byrne & Bates, 2007; de Waal, 2013; Dunbar & Shultz, 2007; Marshall-Pescini, Virányi, & Range, 2015). In particular, the inhibition of feeding and other behaviours when more dominant individuals are present is a common challenge for gregarious animals living in societies structured by dominance hierarchies (Amici et al., 2018). Skills such as tactical deception require a high degree of inhibitory control, and such skills are particularly adaptive in societies where dominance hierarchies determine access to food and mates (Amici et al., 2018; Menzel, 1974; Whiten & Byrne, 1988). For instance, a low-ranking male primate might inhibit his vocalizing during mating to avoid attracting the attention of more dominant males. Indeed, research on primates suggests that living in social systems with steeper dominance hierarchies and higher degrees of fission-fusion dynamics may be associated with better inhibitory control (Amici et al., 2008; Amici, Call, & Aureli, 2009). However, research on the SIH has largely used the comparative approach, which ignores the large amount of observed intra- specific variation in both social demands and cognitive abilities. Recently, there has been growing interest in examining this intraspecific variation in cognitive abilities in wild subjects in order to directly quantify social complexity and cognition in an ecologically valid context (Ashton, Thornton, et al., 2018; Whiten, 2018). Here, we experimentally measured inhibitory control in a population of free-living spotted hyaenas, *Crocuta crocuta*, to test predictions of the SIH in regard to inhibition and inquire how varying social demands might shape variation in inhibitory control.

Spotted hyaenas live in complex social systems that exhibit a high degree of fission-fusion dynamics (Smith, Kolowski, Graham, Dawes, & Holekamp, 2008) and that are structured by strict linear dominance hierarchies that show remarkable convergence with those structuring the societies of many cercopithicine primates (Holekamp et al., 2007). This convergent social complexity makes them ideal subjects for testing hypotheses about the evolution of social intelligence. Spotted hyaena social groups, called clans, range in size from 15 to 130 individuals (Holekamp, Dantzer, Stricker, Shaw Yoshida, & Benson-Amram, 2015) and are composed of multiple matrilines of unrelated females, their offspring and one to several adult immigrant males. Rank is not genetically determined; instead, it is ‘inherited’ behaviourally via maternal interventions and other learning situations, and offspring of both sexes typically rank just below their mothers in the clan's dominance hierarchy (Engh et al., 2000). Female hyaenas are philopatric whereas males nearly always disperse after reaching sexual maturity (2-5 years of age) to join neighbouring clans (Van Horn et al., 2003). Maintenance of rank relationships depends heavily on social support (Strauss & Holekamp, 2019; Vullioud et al., 2019), so the highest- ranking individuals in a hyaena clan are always female because male hyaenas, who lack support from their kin in their new groups after dispersal, join their new clan at the very bottom of its dominance hierarchy. Many male hyaenas must therefore go through the sudden transition of having a relatively high rank at birth in their natal clan to an extremely low rank in adulthood in the clans to which they disperse. Previous research on spotted hyaena cognition suggests that they share many of the same socio-cognitive abilities possessed by cercopithecine primates; for example, spotted hyaenas show individual recognition of clanmates through multiple modalities and the ability to recognize third-party relationships, and are able to use their knowledge about their clan- mates' rank to make adaptive decisions (Holekamp et al., 2007).

The SIH generates several predictions about inhibitory control that we tested in wild spotted hyaenas. Because low-ranking hyaenas must more frequently inhibit feeding and aggression in the presence of higher-ranking individuals than do their higher-ranked groupmates, and because male hyaenas nearly always attain very low ranks as adults, we predicted that adult male hyaenas would have greater inhibitory control than adult females. This prediction was also based on previous research showing that the volume of the frontal cortex is significantly larger in male spotted hyaenas than in female hyaenas; frontal cortex is an area of the brain thought to be centrally involved in social cognition and inhibitory control (Arsznov, Lundrigan, Holekamp, & Sakai, 2010). Next, if inhibitory control is a plastic trait, rather than a fixed one, low-ranking individuals of either sex should also show higher inhibitory control in standardized tests than high-ranking individuals. In addition, we might expect dispersal status of males to play a role: low-ranking immigrant males should have greater inhibitory control than adult natal male hyaenas, of similar age, who have not yet dispersed and who therefore retain their mothers' ranks in the natal group.

In addition to the demands imposed by social rank, larger numbers of individuals in a group may also demand greater inhibitory control (Ashton, Ridley, et al., 2018). Group size is frequently used as measure of social complexity; larger groups may be more cognitively demanding due to increasingly complex social structures and exponential increase in the numbers of relationships and interactions with groupmates (Kappeler, 2019). We therefore predicted that clan size would be positively related to inhibitory control. Finally, most cognitive abilities also show developmental effects; mature individuals typically show enhanced abilities relative to those found in juveniles or subadults (Diamond, 1990, 2013). Therefore, we also predicted that subadult hyaenas would have poorer inhibitory control than adult hyaenas.

METHODS

Subjects and subject participation

Subjects were from two populations of wild spotted hyaenas living in the Maasai Mara National Reserve (MMNR), Kenya. The data used in this study were collected between June 2016 and December 2017, but these two populations have been monitored continuously for demographic and behavioural data from 1988 and 2008 onward, respectively. Thus, all hyaena ages and social rank positions were known. All hyaenas in these populations are well habituated to the presence of observers in research vehicles and are identifiable individually by their unique spot patterns and ear damage, making it possible to collect detailed information on individuals across their life spans. We selected hyaenas from five different clans containing 66e118 individuals per clan during the study period. Unrelated hyaena cubs within the same clan are raised together at a communal den for protection (Holekamp & Dloniak, 2010). Only hyaenas that were independent of the communal den were used for testing, and within that group, hyaenas self-selected as those choosing to participate in the inhibitory control test. Although self-selection is a source of bias, participation was relatively high and a wide representation of all age, sex and rank classes was obtained. Overall, roughly half of all hyaenas given opportunities to participate in a familiarization trial with the inhibitory control apparatus prior to testing chose to do so (92 out of 195 den-independent individuals). Participation was defined as the completion of a trial by feeding from the apparatus. Sixty-two hyaenas subsequently participated in at least one test trial after passing familiarization trial criteria.

Experimental apparatus

We measured inhibitory control using the ‘cylinder task’, which is a scaled-down detour-reaching test (Figure 4.1). Detour-reaching tests in general, and the cylinder task in particular, are used extensively for measuring sensory-motor inhibitory control (Kabadayi et al., 2018). These tests require the subject to detour around a barrier (such as a fence or transparent wall) to reach a visible reward. Detouring requires the subject to initially move away from the reward and inhibit an impulse to move straight towards the reward. A successful trial, or a ‘pass’, is scored if the subject detours without bumping into or touching the barrier, and an unsuccessful trial or a ‘failure’ of inhibition is scored if the subject bumps into or touches the barrier in an attempt to move towards the reward. Detour tests have high ecological validity because animals must detour around obstacles in nature, and high construct validity because they have been long established as accurately measuring executive control (Kabadayi et al., 2018).

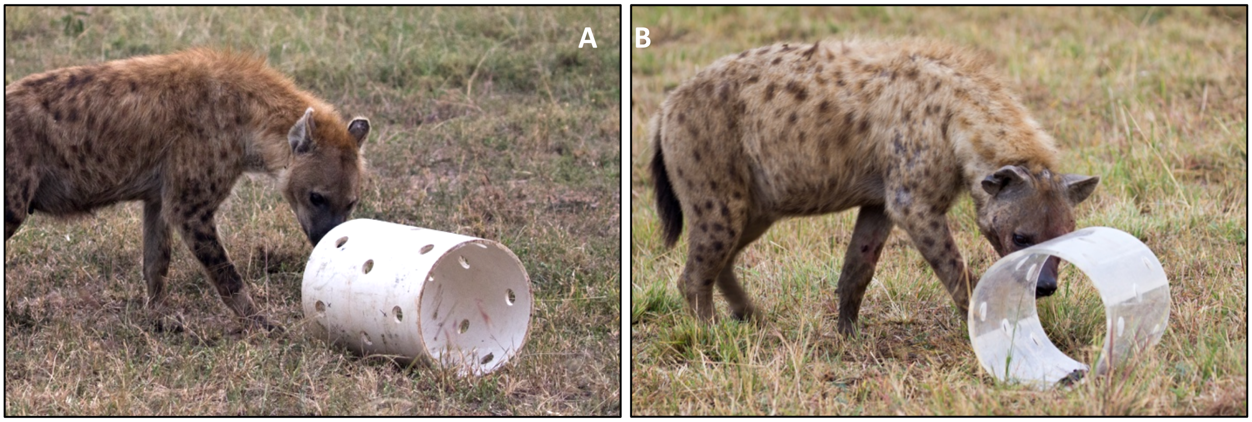

Figure 4.1. The apparatuses used to test inhibitory control; **A)** the white cylinder used during familiarization trials and **B)** the transparent cylinder used during test trials.

We also thought a detour task would be particularly appropriate for testing inhibitory control in spotted hyaenas because it specifically measures the ability to inhibit approach towards a food reward, just as low-ranking hyaenas must inhibit approaching food when a higher-ranking hyaena is present. The cylinder detour task is also well suited for testing in the field because its apparatus is easy to deploy; we have also used this task successfully with captive hyaenas(Johnson-Ulrich et al., 2018). The relatively small size of the cylinder means that it does not create an artificial requirement for spatial reasoning, and unlike non-detour tests of inhibitory control, it requires no knowledge of object permanence or other object affordances (Jelbert, Taylor, & Gray, 2016). Variation in experience with clear objects can influence subjects' performance with the cylinder task (van Horik et al., 2018), but all subjects in the current study were wild hyaenas that were all completely inexperienced with clear objects. The cylinders used here measured 30.5 cm in diameter and 46 cm in length and were constructed of1 cm thick plastic. A white PVC cylinder was used for familiarization trials (Figure 4.1A), and a clear cast acrylic cylinder (Figure 4.1B) was used for test trials. Spotted hyaenas have a keen sense of smell, so unlike previous studies with the cylinder task, we bored 15 holes (2.5 cm diameter) into the wall of each cylinder to control for the fact that the scent of a food reward might lead the hyaenas to successfully retrieve food without inhibiting their approach to the food directly.

Test protocol

Focal hyaenas were located during daily observation periods from 0530 to 0930 hours and again from 1700 to 2000 hours, which are the periods each day when spotted hyaenas are most active. Hyaenas were only tested when they were alone, which was determined if no other hyaenas were in sight or known to be present within approximately 200 m. The cylinder was baited with a piece of meat weighing approximately 200 g, placed directly in the centre of the cylinder. Bait was nearly always goat meat, but some trials were conducted with beef or mutton. Participation in trials with the baited apparatus appeared unaffected by the type of bait used (Table 4.A.1). Our research vehicle served as a mobile blind from which the cylinder was placed horizontally on the ground with the long side facing the hyaena. After deployment, the research vehicle was driven 20-50 m away, and the entire trial was videotaped. Trials began when the cylinder was deployed. Trials ended when either the bait was retrieved, or when the focal hyaena failed to participate by going of sight, moving outside the 200 m test radius, or lying down for at least 15 min. Trials where hyaenas failed to participate were not included in our analysis. Average (± SD) trial duration was 2.85 ± 2.76 min (median = 2 min, range = 0 – 23 min). If other hyaenas arrived during a trial, we collected the still-baited apparatus and paused testing until the other hyaena(s) moved to more than approximately 200 m from the focal hyaena or went out of sight. If the newly arrived hyaena(s) did not leave within 20 min, we ended testing with the focal hyaena for the day. All hyaenas were initially given familiarization trials using the opaque white PVC cylinder. A 'pass' was recorded if the hyaena successfully retrieved the meat without touching the outside of the cylinder. A 'fail' was recorded if the hyaena touched the outside of the cylinder with its nose. Hyaenas sometimes sniffed the outer rim of the cylinder before they inserted their heads inside the cylinder to retrieve the food, especially during familiarization trials; this touching of the outer rim only was considered exploratory behaviour (see Kabadayi et al., 2017). Therefore, touching just the outer rim, which was immediately before the hyaena inserted its head into the cylinder to retrieve the meat, was not recorded as a fail. After a hyaena passed four out of five consecutive familiarization trials (with a minimum of five trials), it was given a maximum of 11 test trials with the transparent acrylic cylinder. While all attempts were made to give every hyaena at least 10 test trials, the difficulties of field testing meant that not all hyaenas received this number of trials (mean ± SD = 8.92 ± 3.09 trials, median = 10 trials, range = 1 – 11 trials). Only 12 out of 60 hyaenas received fewer than 10 trials. If a hyaena retrieved and ate the bait, and if it had moved at least 5 m away from the cylinder, the cylinder was immediately rebaited for successive trials. Successive trials were administered to a hyaena as long as conditions allowed researchers to do so. For example, successive testing was discontinued if the hyaena stopped participating (as described above), bait supply was exhausted, or rain, other inclement weather or darkness made testing impracticable. Mean (± SD) test session duration was 10.62 ± 10.63 min (median = 8 min, range = 0 – 57 min). Hyaenas were given an average (± SD) of 4.64 ± 2.96 trials per session (median = 4 trials, range = 1 – 10 trials) and completed testing in 1.92 ± 0.96 sessions (median = 2 sessions, range = 1 – 5 sessions). Most test trials were completed less than 1 day apart (N = 479 out of 539 trials); however, some hyaenas were not located again for further testing for extended periods. The average (± SD) number of days between trials was 5.32 ± 30.74 days (median = 0 days, range = 0 – 280 days). Only 10 trials were conducted with more than 100 days separating them from the last trial with the same hyaena and these 10 trials did not appear to differ in any measurable way from all other trials. Specifically, removing these 10 trials from our analyses did not alter the significance of any independent variables in our final models. Because hyaenas were given 360-degree access to the tube, we also recorded the direction of their approach as perpendicular, parallel or ‘other’ if the approach did not follow a straight line. In addition, hyaenas approached the cylinder at variable speeds, potentially indicating variation in motivation. Therefore, we recorded their latency to approach from 5 m to 1 m of the cylinder in seconds. Both latency to approach and direction of approach were included as extraneous variables in our statistical analysis. All videos were coded by L.J.U.; videos were named with a unique trial number.

Predictor variables and extraneous variables

We had five main predictions about intraspecific variation in inhibitory control regarding social rank, sex, dispersal status, clan size and age of spotted hyaenas. Social rank was treated as a continuous variable; all individuals in a clan were assigned a numeric rank based on their position in the hierarchy. Ranks used in the current analysis were assigned based on the clan hierarchy at the start of the study in June 2016. This hierarchy was generated using a dominance matrix ordering observations of aggressive or submissive behaviours within dyads of hyaenas over the age of 24 months (Holekamp, Smith, Strelioff, Van Horn, & Watts, 2012; Martin & Bateson, 1993). Hyaenas younger than 24 months of age were assigned a rank directly below that of their mother. Although an individual hyaena's rank may vary across time, most rank changes are small and caused by births and deaths such that rank relative to other clan members remains stable even if the numeric value fluctuates slightly (Strauss & Holekamp, 2019). Rank reversals are rare in this species (Vullioud et al., 2019), and none are known to have occurred during the study period. In our analysis, each hyaena's numeric rank was scaled and centred, with a score of 1 for the highest-ranking individual in a clan and a score of -1 for the lowest. Scaling and centering was done in order to obtain a relative, rather than absolute, value for social rank that was independent of clan size. Hyaenas were sexed based on the morphology of the penis in males or pseudo-penis in females. Dispersal status was described either as ‘natal’, for hyaenas born in the current clan, or as ‘immigrant’, for immigrant males. Group size was measured in two different ways. First, overall clan size was approximated as the total number of individuals in a clan at the time of testing. Group size is also known to have effects on cognition during early life development (Ashton, Ridley, et al., 2018); we therefore also measured group size as the number of individuals in each hyaena's communal den cohort. The size of the communal den cohort was calculated as the number of cubs whose period of residence at the communal den overlapped for at least 4 months with that of every other young hyaena present at the den. In our sample, sizes of these communal den cohorts ranged from 4 to 31 cubs. We initially calculated age as a continuous variable rounded to the nearest month; however, age appeared to have a nonlinear relationship with inhibitory control (see Statistical Analysis) and it was therefore binned into three categories: subadults (0-2 years), young adults (2-5 years) and adults (>5 years).

We also measured several extraneous variables that were related to the conditions of each trial, other characteristics of each hyaena subject, or characteristics of the testing environment that might have affected trial success independent of a hyaena's inhibitory control. Hyaena body condition was included with three levels (gaunt, unremarkable or fat); body condition generally indicates how recently a hyaena has eaten (Kruuk, 1972) and might affect motivation to obtain a food reward. Likewise, the annual wildebeest migration in the MMNR represents a massive influx of prey for spotted hyaenas (Holekamp et al., 1997) and this too could potentially affect motivation. Therefore, we also included ‘season/migration’ to indicate whether or not the migration was present in the Reserve when a particular trial was conducted. Although every possible attempt was made to test hyaenas only when they were alone, during 48 of 539 trials, one or two other individuals also showed up. Because this number of additional hyaenas present only varied from one to two, we included testing group size as a categorical variable of either one or more than one. We also included a variable to indicate the number of these additional hyaenas that were higher ranking than the hyaena being tested, and this variable was likewise binned as a categorical variable of either one or more than one additional higher-ranking hyaena. Some studies using the cylinder task have reported an effect of learning across trials, which can positively bias results (Kabadayi et al., 2017). To control for any potential learning effects, we recorded trial number both across all trials and within each testing session. Because the amount of time between trials and between testing sessions varied, we also included a variable for the number of days elapsed since a hyaena's last trial (see Test Protocol). The number of days elapsed since a hyaena's last trial also appeared to have a nonlinear relationship with inhibitory control; because nearly all trials occurred on the same day, we binned this variable into a binary variable of zero days and one or more days. Last, we included variables for the latency to approach and direction of approach, as described above in Test Protocol. Latency to approach was binned into a categorical variable of either fast or slow because it appeared to have a nonlinear relationship with inhibitory control and was highly left-skewed. The median latency to approach was 6 s (mean ± SD = 11.96 ± 23.62), so we classified fast approaches as those in which the hyaena's latency to get from 5 m to within 1 m of the cylinder was less than or equal to 6 s, and slow approaches as those in which the hyaena's latency was greater than 6 s.

Statistical analysis

We analysed the factors affecting inhibitory control using generalized linear mixed models (GLMMs) with a logarithmic link function. All statistical analysis was done using R v.3.5.0 (R Core Team, 2019). Inhibitory control, our dependent variable, was recorded binarily as 'pass' or 'fail'. One assumption of binomial models is that independent variables show a linear relationship with the dependent variable on a log-odds scale (Burnham & Anderson, 2002). Therefore, prior to the creation of any model, all continuous variables were examined to determine whether they met this assumption. Integer count variables with less than four unique values were treated as categorical. This was done for age, clan size, test group size and number of days passing between a hyaena's trials (see Predictor Variables and Extraneous Variables). We also examined evidence for multicollinearity using variance inflation factors (VIFs) prior to creating our global models. Any variable with a VIF >3 was excluded (Zuur, Ieno, & Elphick, 2010). Ultimately, all VIFs were <3.

We created three models, the first to test predictions about sex, rank, age effects and clan size effects (N = 60 subjects), the second included data only from male hyaenas to test predictions about dispersal status (N = 26 subjects), and the third included only data from hyaenas for whom we were able to calculate cohort size (N = 42 subjects) to test predictions about effects of this form of group size. All three models included rank, age and clan size. Model 2 did not include sex because only males were included in this model, but it did include a binary natal versus immigrant variable. Model 3 added the variable cohort size. All three models included a random effect of hyaena identity.

To investigate the effect of extraneous variables on inhibitory control, we also created three 'global' models that added every extraneous variable we measured to our initial three models (see Predictor Variables and Extraneous Variables above). Because we had no a priori expectations about the size or direction of the effect of these extraneous variables, we initially used exploratory model analysis to identify which extraneous variables were actually important (Whittingham, Stephens, Bradbury, & Freckleton, 2006). For exploratory model analysis, we used the Dredge function in the R package MuMIn (Bartoń, 2018). Dredge is a model selection function that systematically examined models containing all possible combinations of the extraneous variables included in our three global models. Variables for which we had specific predictions (e.g. rank and sex) were always included in every variable combination examined across models. Models were ranked by Dredge for their goodness of fit using Akaike's information criterion with a correction for small sample sizes (AICc). Top models were defined as those for which the improvement in AICc was <4 (Burnham & Anderson, 2002; Wisnieski et al., 2019). Although Dredge did not produce a single top model from any of our three global models, only two extraneous variables, latency to approach and direction of approach, were included in all top models for each of our three global models (Table 4.A.2). These were also the only two variables to have large and significant effects in every top model. None of our other extraneous variables had significant effects. Therefore, we decided to include only latency to approach and direction of approach in our three final models along with the variables for which we had specific predictions. Finally, we assessed the repeatability of successful trials across individual hyaenas using the rptR package (Stoffel et al., 2017). We calculated repeatability for hyaena identity (ID) using a null model that included only the trial number as a fixed effect and hyaena ID as a random effect with a binary response variable (pass/fail).

RESULTS

We conducted 539 test trials with 62 hyaenas in five different clans. Two hyaenas were of unknown sex and their trials were dropped from the analysis, resulting in a final data set of 529 trials involving 60 hyaenas. Within these 60 hyaenas, ages ranged from 7 to 188 months. These hyaenas included 34 female and 26 male hyaenas (20 subadults,19 young adults and 24 adults). Five hyaenas moved from one age class to the next during testing, and their trials were assigned to their age class at the time of each trial. Overall, the mean (± SD) proportion of correct responses with the opaque cylinder used in familiarization trials by hyaenas was 0.85 ± 0.15 (median = 0.8, range = 0.33 – 1, N = 60) and the mean proportion of correct responses with the transparent cylinder used in test trials by hyaenas was 0.76 ± 0.21 (median = 0.8, range 0 – 1, N = 60). On average, hyaenas had nearly one fewer successful trial with the clear tube than with the white out of ten trials (paired two-tailed *t*-test: $\text{mean}_{\text{diff}} = 0.09$, $t_{59} = -2.62$, $P = 0.01$). This suggests that the transparent cylinder presented more of an inhibitory challenge for hyaenas than the opaque tube. Repeatability of success on individual trials with the transparent cylinder was significant but low (Likelihood ratio test: R = 0.07, $P = 0.001$).

We created three models to examine the relationship between sex, age, rank, clan size (Model 1), immigration status (Model 2) and cohort size (Model 3) on inhibitory control. We included two extraneous variables, latency to approach and direction of approach, and one random effect of hyaena ID in all three models. We also added an interaction between rank and clan size to all three models even though we had no a priori predictions about an interaction between these two variables. We added this variable post hoc because in Model 2 (male subset) clan size was initially strongly significant, whereas in Models 1 and 3 it was not. The average rank of hyaenas in our male subset was somewhat lower than the overall average because immigrant males are always very low ranking. This led us to suspect that there might be an interaction between clan size and rank, and this interaction was subsequently added to all three models. We conducted residual diagnostics and evaluated model fit using the R package Dharma (Hartig, 2019). We examined fit by plotting the scaled residuals, a residual dispersion test and the Durbin-Watson test for temporal autocorrelation in the residuals. Model fit was good for all three final models (Figures 4.A.1-4.A.3). Scaled residuals did not significantly deviate from expected values and their distribution was generally uniform without skew. There was no significant dispersion or temporal autocorrelation in the residuals. The results from these tests indicated that model fit was adequate.

Table 4.1. Model results for three final models looking at the factors affecting inhibitory control.

	Model 1: Full				**Model 2: Male subjects**				**Model 3: Natal subjects**			
	Odds ratio	**SE**	*z*	*P*	**Odds ratio**	**SE**	*z*	*P*	**Odds ratio**	**SE**	*z*	*P*
Sex (male)	1.06	0.28	0.21	0.833	-	-	-	-	1.40	0.31	1.10	0.273
Rank	0.87	0.26	-0.52	0.603	0.57	0.83	-0.68	0.497	0.92	0.28	-0.31	0.756
Age (sub)	1.41	0.33	1.03	0.305	0.36	1.16	-0.87	0.383	1.44	0.36	1.02	0.309
Age (young adult)	1.83	0.32	1.91	0.056	1.07	0.89	0.08	0.935	1.84	0.34	1.78	0.075
Clan size	1.32	0.14	2.04	0.041*	1.74	0.31	1.77	0.077*	1.06	0.18	0.30	0.762
Latency	0.47	0.26	-2.90	0.004*	0.61	0.44	-1.12	0.261	0.53	0.31	-2.10	0.036*
Approach (other)	2.93	0.31	3.43	0.001*	2.56	0.53	1.76	0.078	3.68	0.37	3.52	0.0004*
Approach (parallel)	6.82	0.75	2.55	0.011*	9.43	1.10	2.04	0.041*	3.70	0.77	1.70	0.090
Rank x clan size	0.52	0.23	-2.82	0.005*	0.54	0.54	-1.16	0.246	0.67	0.27	-1.50	0.133
Immigrant	-	-	-	-	0.26	1.48	-0.91	0.363	-	-	-	-
Cohort size	-	-	-	-	-	-	-	-	1.05	0.02	2.09	0.036*

Note: All models were built using binomial generalized linear mixed models with a log link function and included random effect of hyena ID. Model 1 used our full dataset (n = 529, N = 60), model 2 used only male hyenas (n = 230, N = 26) and model 3 used only hyenas for which we had natal data (n = 384, N = 42). The reference categories were adult, female, and a perpendicular approach for sex, age, and approach.
*Indicates statistical significance at α = 0.05

Model 1 investigated factors influencing inhibitory control using our full data set (Table 4.1: Model 1). We found no effect of sex on the odds of a successful trial (GLMM: $z = 0.21$, $P =$

0.83). Young adults had slightly higher odds of success than older or younger hyaenas, but this effect was marginally nonsignificant (GLMM: $z = 1.91$, $P = 0.056$; Figure 4.2).

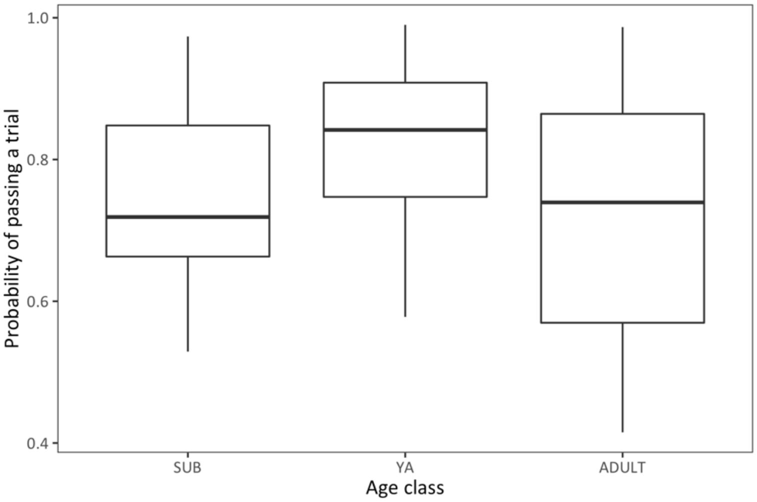

Figure 4.2. Predicted probability of success for subadult (N = 20), young adult (N = 19) and adult (N = 24) spotted hyaenas estimated from Model 1. Box plots show median values and interquartile ranges.

Hyaenas who had shorter latencies to approach the cylinder had much lower odds of success (GLMM: $z = -2.90$, $P = 0.004$), whereas hyaenas who used an ‘other’ approach, such as diagonal or circling approach, had higher odds of success (GLMM: $z = 3.43$, $P = 0.001$). Hyaenas living in larger clans had modestly but significantly better odds of success than those living in smaller clans (GLMM: $z = 2.04$, $P = 0.04$; Figure 4.3). Finally, although rank alone was not significant (GLMM: $z = -0.52$, $P = 0.60$; Table 4.1: Model 1), we found a significant interaction between rank and clan size (GLMM: $z = -2.82$, $P = 0.005$). At a very low social rank, the coefficient for clan size was large and significant, which shows that for low-ranking hyaenas, living in larger clans was correlated with higher odds of a successful trial and greater inhibitory control (Figure 4.4).

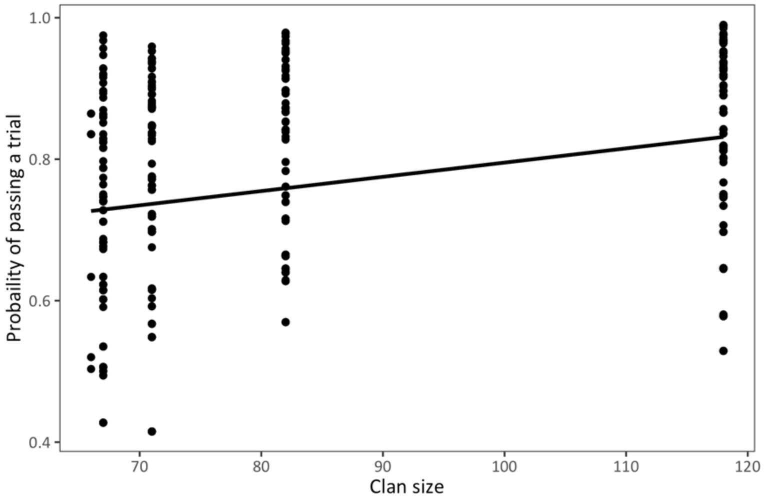

Figure 4.3. Predicted probability of success for hyaenas living in clans of different sizes estimated from Model 1.

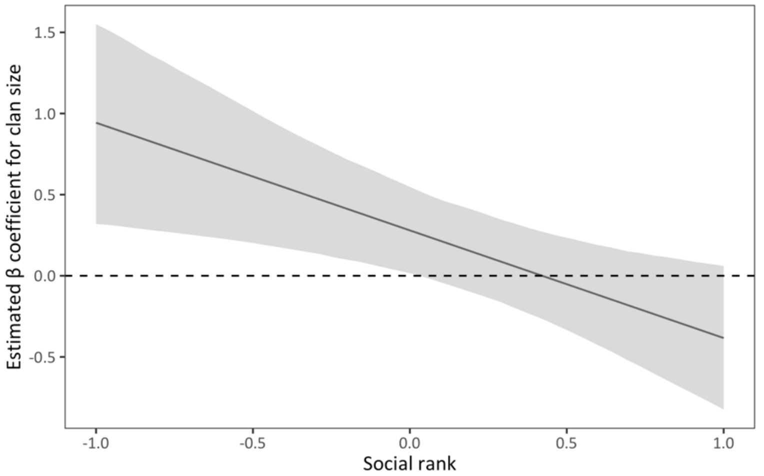

Figure 4.4. Interaction plot between rank and clan size. Figure shows the effect of rank on the estimated coefficient (log odds) of clan size estimated from Model 1.

However, as social rank increased, the effect of clan size became weaker and nonsignificant, such that for high-ranking hyaenas, the size of the clan had little effect on their odds of success. There also appeared to be no effect of presentation order or learning across all trials or across trials within a testing session. Trial number was included in only six out of 18 top models for Model 1 produced by Dredge (GLMM conditional average: $z = 0.91$, $P = 0.37$; Table

4.A.2, Figure 4.A.4). Session trial number was also included in only six out of 18 top models for Model 1 produced by Dredge (GLMM conditional average: $z = 0.72$, $P = 0.47$, Table 4.A.2, Figure 4.A.5).

Model 2 investigated the effect of immigration status on inhibitory control in a subset of subjects that included only males (Table 4.1: Model 2). Immigrant males did not have significantly higher odds of success than natal males (GLMM: $z = -0.91$, $P = 0.36$).

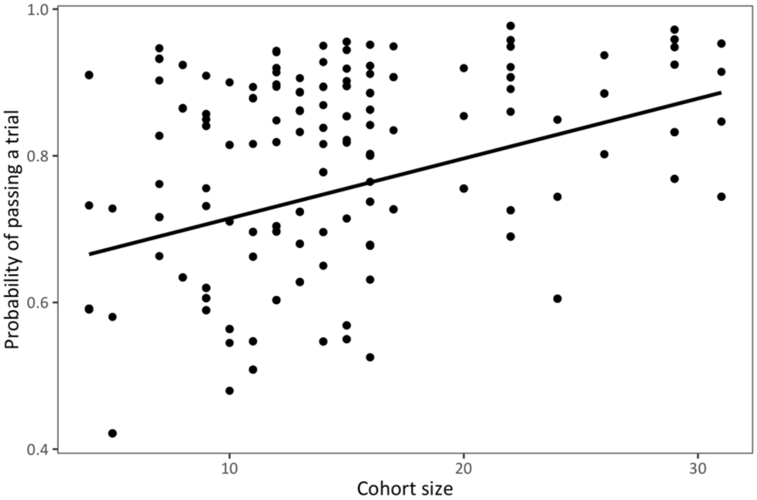

Figure 4.5. Predicted probability of success for hyaenas growing up in den cohorts of different sizes estimated from Model 3.

Model 3 investigated the effect of cohort size on inhibitory control in a subset that included natal animals for which cohort size was known (Table 4.1: Model 3). In this model cohort size had a significant positive effect on the odds of a successful trial (GLMM: $z = 2.09$, $P = 0.036$). Hyaenas that grew up in larger cohorts of cubs at the communal den had significantly better odds of success than those that grew up in smaller cohorts (Figure 4.5).

DISCUSSION

Testing the social intelligence hypothesis

This study examined intraspecific variation in inhibitory control in wild spotted hyaenas in relation to social factors in order to test predictions of the social intelligence hypothesis (SIH) (Dunbar, 1998; Dunbar & Shultz, 2017; Humphrey, 1976). Although the SIH is one of the most popular hypotheses forwarded to explain the evolution of intelligence in animals, it is still somewhat controversial, with some interspecific comparative studies finding strong support and others finding no support at all for the SIH (Barrett, Henzi, & Rendall, 2007; DeCasien et al., 2017; Fedorova, Evans, & Byrne, 2017; Powell et al., 2017; Rosati, 2017). In particular, how to define and quantify social complexity is still debated (Bergman & Beehner, 2015; Boucherie, Loretto, Massen, & Bugnyar, 2019; Kappeler, 2019). Here, we attempted to directly quantify intraspecific variation in social demands on cognition and compare this variation to variation in inhibitory control, an executive function thought to be particularly important in gregarious species. Although we found that inhibitory control did not vary with sex, both clan size and cohort size, which was the size of the group of peers in which hyaenas spent their early life, were significant predictors of inhibitory control, with hyaenas in larger clans (Figure 4.3) and hyaenas raised in larger cohorts (Figure 4.5) exhibiting greater inhibitory control. In addition, we found a significant interaction between rank and clan size, with low-ranking hyaenas exhibiting higher levels of inhibitory control in larger clans (Figure 4.4). Group size has frequently been used as a measure of social complexity in interspecific studies. Spotted hyaenas live in large groups relative to those of other social carnivores, but in groups of similar size to those of cercopithicine primates. Importantly, there is a great deal of intraspecific variation in group size within spotted hyaenas, with clans ranging in size from 22 to 126 individuals in the MMNR (Green, Johnson-

Ulrich, Couraud, & Holekamp, 2018). Comparing this variation in clan size to variation in cognition within the species provides a powerful tool for directly testing the effect of group size on cognition. In the present, study clan sizes ranged from 66 to 118 total members (Figure 4.3) and clan size was significantly and positively correlated with inhibitory control (Table 4.1, Model 1). Social relationships during development may also be of particular importance in social animals (Ashton, Ridley, et al., 2018). Until approximately 1 year of age, spotted hyaenas live at a communal den with other unrelated juvenile spotted hyaenas, and it is here that cubs begin to learn their dominance ranks and acquire crucial social and physical skills (Holekamp & Dloniak, 2010; HOLEKAMP & SMALE, 1991). We therefore also tested the effect of cohort size in the subset of our subjects for which cohort size was known. In this model, cohort size strongly and positively predicted inhibitory control, with hyaenas that grew up in larger cohorts of cubs possessing higher levels of inhibitory control. Interestingly, cohort seemed to provide a stronger fit than clan size; when cohort size was added to Model 4.3, clan size became nonsignificant (Table 4.1: Model 3). This suggests that the effect of group size may be most important during early development, when both social and cognitive skills are still developing. A study in Australian magpies, *Gymnorhina tibicen*, where general intelligence was measured by a cognition test battery that included the cylinder test, also found evidence for a relationship between group size and general intelligence that emerged very early in development (Ashton, Ridley, et al., 2018). Our results replicate and support the findings of Ashton et al. in a mammalian carnivore, greatly strengthening the argument for a relationship between cognitive ability and complexity in the early life social environment. We had also predicted that rank would influence inhibitory control because low-ranking individuals must inhibit feeding from carcasses whenever higher-ranking hyaenas are present, but instead found that the effect of rank was contingent on clan size. While unpredicted, this result is not entirely surprising. Scaling rank relative to the other members of a group is standard practice, but it does not necessarily capture the difference in social demands between a low-ranking hyaena in a small clan and a low-ranking hyaena in a large clan. In a small clan, a low-ranking hyaena might rank below only 20 other hyaenas, whereas in a large clan, a low-ranking hyaena may rank below 120 other hyaenas. The significant interaction between rank and clan size suggests that low- ranking hyaenas in larger clans may experience greater demands for better inhibitory control, perhaps due to the larger number of individuals ranked above them. Our findings suggest that both the size and social structure of a group are important aspects of social complexity in spotted hyaenas. Therefore, this result not only supports previous work reporting a relationship between group size and cognition but also expands our understanding of how and why variation in group size can affect demands on cognitive abilities. Last, we also predicted that male hyaenas, which are nearly always extremely low ranking as adults, and which have larger frontal cortices than females (Arsznov et al., 2010), would have the highest levels of inhibitory control. However, sex was not significant in our final model (Table 4.1: Model 1). It is worth noting that, although the size of the frontal cortex varies between male and female hyaenas (Arsznov et al., 2010) and although the relative size of the entire neocortex has consistently been linked to social intelligence (reviewed in Holekamp & Benson-Amram, 2017), previous researchers using the cylinder task found that performance was more closely correlated with overall brain size than with frontal cortex or neocortex size alone (Horschler et al., 2019; MacLean et al., 2014). The lack of significant sex differences in this study suggests that inhibitory control may not be a canalized trait and may instead be a plastic trait contingent on developmental conditions (e.g. the social environment). Because rank, cohort size and clan size, none of which are genetically determined, are all strongly related to inhibitory control in hyaenas, our findings also support this notion that inhibitory control is a plastic trait.

Taken together, our results provide strong support for the social intelligence hypothesis; by demonstrating that both measures of social group size and a rank by clan size interaction predict inhibitory control, this study provides direct support for the hypothesis that social complexity favours the development of enhanced cognitive abilities. The size of the group and an individual's social rank within it both appear to place cognitive demands on hyaenas, favouring the development of greater inhibitory control. Direct evidence for such a relationship between greater social complexity and enhanced cognition has been surprisingly rare. Most support for the SIH comes from studies that use average group size and average brain size of each species, rather than direct measures of cognition, across various species in comparative analyses (Dunbar, 1998; Dunbar & Shultz, 2007). Interestingly, interspecific studies of captive animals in which cognitive ability was directly measured have generally failed to find a link between cognitive performance and group size (Benson-Amram et al., 2016; MacLean et al., 2014; Reddy, MacLean, Sandel, & Hare, 2015; Rudolph & Fichtel, 2017). Indeed, it is still unclear just how or why a larger group increases cognitive demands, although the SIH offers many suggestions (Aureli & Schino, 2019; Barrett, Henzi, & Dunbar, 2003; Boucherie et al., 2019; Emery, Seed, von Bayern, & Clayton, 2007; Kappeler, 2019). For example, it may be that managing a greater number of differentiated social relationships requires greater intelligence (Bergman & Beehner, 2015), but it might also be that living in a larger group exposes individuals to more frequent opportunities to learn socially about their physical environment; under these circumstances social living is a facilitator rather than a selective force for intelligence (van Schaik et al., 2012) as proposed in the cultural intelligence hypothesis (CIH). Although we are currently unable to differentiate between these two hypotheses, captive spotted hyaenas show only very limited social learning of physical skills (Benson-Amram et al., 2014). In addition, previous research with spotted hyaenas found that rates of social play are related to cohort size and that rates of social play, but not object play, peak while cubs live at the communal den (Tanner, Smale, & Holekamp, 2007). Because social play is thought to be critical to development of social skills in hyaenas (Drea, Hawk, & Glickman, 1997), these previous studies certainly suggest that aspects of sociality itself are more likely to directly mediate the relationship between cohort size and inhibitory control, rather than indirectly as the CIH suggests. Indeed, the interaction between clan size and rank found in the present study certainly suggests that navigating the social hierarchy is particularly challenging for low-ranking hyaenas. This finding also supports the idea that managing a large number of differentiated social relationships is a critical aspect of social complexity. Ultimately, more research is needed to test these possibilities. Potential areas for future research include using social network analyses to quantify the strength of social relationships or using experimental paradigms to directly measure social learning and cultural diffusion in wild hyaenas.

Developmental effects on inhibitory control

Many cognitive abilities are developmentally plastic, and individuals tend to show improved cognition as they approach adulthood due to learning and brain development (CASEY, TOTTENHAM, LISTON, & DURSTON, 2005; Greenough, Black, & Wallace, 1993; Johnson, Munakata, & Gilmore, 2005; Kolb & Gibb, 2011; Snell-Rood, 2013). Executive functions, such as inhibitory control, are typically among the last to become fully developed during ontogeny (Diamond, 1990, 2013). We therefore had predicted that adult spotted hyaenas would show better inhibitory control than subadults. However, we found that young adult hyaenas 2-4 years of age had higher levels of inhibitory control than other age groups (*P* = 0.055). This effect was not driven by male hyaenas (which disperse during this age interval), as age was not a significant predictor of inhibitory control in our male-only model. Instead, this effect may simply be a product of the way the brain develops, rather than driven by life history or environmental circumstances. Research on human and animal cognition suggests that many cognitive abilities peak in young adulthood (Craik & Bialystok, 2006; Feyereisen & Charlot, 2008; Matzel, Grossman, Light, Townsend, & Kolata, 2008; Salthouse, 2009; Whitley et al., 2016). Whereas crystallized intelligence, a facet of general intelligence that represents accumulated knowledge across the life span, continues to improve across the entire life span, fluid intelligence, which is akin to behavioural flexibility in the animal literature, tends to peak during young adulthood. However, the exact age at which this peak occurs varies tremendously among different cognitive abilities (Hartshorne & Germine, 2015). Although our sample included only den-independent cubs, which are usually older than 8 months of age, it would be extremely interesting to test young hyaenas at the den to examine how inhibitory control might improve from a very young age.

Inhibitory control task design considerations

The cylinder task is fairly well established as a tool for studying inhibitory control, but performance with the cylinder task has been found to be susceptible to several sources of bias. For one, prior experience with clear objects can significantly improve performance on detour tasks that use clear barriers (van Horik et al., 2018; but see Santacà, Busatta, Savaşçı, Lucon-Xiccato, & Bisazza, 2019) and several studies have also found that subjects tend to improve in their success rate across trials (Kabadayi et al., 2017), suggesting that individuals may be learning the affordances of the clear cylinder, which would make the inhibitory challenge decrease across trials. This can bias results by making it look like individuals have higher levels of inhibitory control than they really do (Kabadayi et al., 2017). Here, all of our subjects were equally inexperienced with transparent objects. To observe and control for any potential learning effects, we recorded trial number both across all trials and within each testing session. Neither of these measures was significant, suggesting that our hyaenas were not learning to perform better with the cylinder across trials. Aspects of test set-up, such as distance and direction of approach can also influence trial success (Kabadayi et al., 2018). Here, hyaenas who had an approach other than perpendicular or parallel to the tube were much more likely to pass a trial. This 'other' category included diagonal and circling approaches, which may have given hyaenas more time to detect the food and locate the openings on either end of the cylinder than more direct approaches, improving their ability to inhibit bumping into the wall of cylinder. Likewise, hyaenas who approached the cylinder more slowly were much more likely to pass a trial. Although latency to and direction of approach may also be measures of a hyaena's inhibitory control, latency to approach is often used as a measure of motivation in studies of animal personality and cognition. Motivation is a known confound in studies of inhibitory control and problem solving (van Horik et al., 2018, 2017), and one study with the cylinder found that temperament and arousal levels affect inhibitory control in domestic dogs (Bray, MacLean, & Hare, 2015). Therefore, we included both of these approach measures in our models to control for their potentially confounding effects. Our results confirm the need to statistically control for such variables, particularly in studies with wild subjects where numerous factors cannot be controlled experimentally. We also controlled for olfaction in our study because, although spotted hyaenas have good visual acuity (Kruuk, 1972), they are highly olfactory animals. Therefore, we drilled holes into the wall of the cylinder so that scent could not draw hyaenas to a successful detour.

Although there is little doubt that detour tasks, like the cylinder test, do measure inhibitory control, inhibitory control on the whole appears to be both task and context specific. Previous research with captive spotted hyaenas did not detect a relationship between inhibition with the cylinder task and inhibition with a multi-access problem-solving box (Johnson-Ulrich et al., 2018). During the cylinder task, individuals must remember their previously learned response with the white cylinder and inhibit an impulse with the clear cylinder, whereas with the multi-access box, individuals are required to inhibit their previously learned solution in order to learn a new one. In general, studies that have used different tasks to measure inhibitory control in animals find that inhibitory control does not correlate across tasks that make different behavioural demands (Bray, Maclean, & Hare, 2014; Brucks, Marshall-Pescini, & Range, 2019; Brucks et al., 2017; Fagnani, Barrera, Carballo, & Bentosela, 2016; Marshall-Pescini et al., 2015; Müller et al., 2016; Völter, Tinklenberg, Call, & Seed, 2018). Indeed, it is likely that inhibitory control is a multifaceted cognitive ability with many underlying processes (Bari & Robbins, 2013). Because of this task specificity, we chose the cylinder test because it most closely approximated the type of inhibitory control we aimed to test in hyaenas: the ability to resist feeding in circumstances that demand restraint (i.e. response inhibition in the context of feeding). It is possible that feeding response inhibition in the presence of a physical barrier, i.e. the cylinder test, might not actually approximate feeding response inhibition in the presence of a 'social barrier' (i.e. a higher-ranking individual), because the costs and benefits of inhibition differs dramatically between these scenarios. This possibility is supported by previous studies showing that social tasks involving inhibitory control do not always correlate with physical tasks involving inhibitory control (Bray et al., 2014; MacLean et al., 2013). However, the inhibitory control tests in these studies not only differed in the physical and social context, but also differed dramatically in task demands. Although we cannot rule out the possibility that our results are biased by using a physical, rather than social, test of inhibitory control, we would expect our results to be biased towards our null hypothesis. For example, if social and physical feeding inhibition were only weakly related or unrelated, we would not have expected to find a significant correlation between our social measures and performance on the cylinder test. However, we did find a significant correlation between our social measures and performance on the cylinder test, which implies that the cylinder test likely does, to some degree, approximate social inhibitory control.

Conclusion

In summary, our results support the SIH as an explanation for the evolution of inhibitory control and provide some of the first direct evidence for a relationship between the social complexity experienced early in life and cognitive ability later in life. We compared two measures of social complexity, social rank and group size, to inhibitory control. We found no support for the hypothesis that male spotted hyaenas, which must frequently inhibit behaviours in the presence of higher-ranking individuals, have better inhibitory control than female hyaenas. Rather, spotted hyaenas who grew up in larger cohorts, lived in larger clans and were low ranking in larger clans had significantly better inhibitory control than other hyaenas. Our results suggest that both the dominance structure and the size of the group experienced during development may be key aspects of social complexity. We suggest that future research should further investigate the mechanism by which group size and structure are related to enhanced cognition. We also found significant effects of age, latency to approach and direction of approach on inhibitory control. Future work should thus further investigate the effects of context and task demands on inhibitory control. It would be extremely interesting to compare a measure of inhibitory control in a social context to both performance on the cylinder test and measures of social complexity. Overall, testing cognition in the wild poses unique challenges for controlling a wide variety of extraneous variables, but the intraspecific approach in wild animals also appears to be a powerful tool for testing hypotheses about the evolution of cognition.

APPENDIX

Table 4.A.1. Results of a binomial regression model on participation by hyaenas in trials with baited apparatus.

	Odds ratio	**SE**	***z***	***P***
Bait (milk powder)	1.77	0.14	3.91	<0.001*
Bait (bone)	0.89	0.12	-1.06	0.29
Bait (meat)	0.96	0.16	-0.22	0.83
Bait (offal)	0.83	0.19	-1.05	0.30
Bait (rotten)	1.02	0.14	0.16	0.87

Note: A score of 1 indicated a hyaena participated in a trial by contacting the apparatus and/or feeding from inside the apparatus and a score of 0 indicated a hyaena failed to participate in a trial. Model included a random effect of hyaena ID (n = 2909 trials, N = 300 subjects). Note that 'milk powder' was not used as bait in cylinder trials.
*Indicates statistical significance at α = 0.05

Table 4.A.2. Dredge results showing the number of top models, determined by a deltaAICc of less than 4, containing each variable.

	Model 1	**Model 2**	**Model 3**
Total Top Models	**18**	**41**	**130**
Sex	18	-	130
Rank	18	-	130
Age class	18	41	130
Clan size	18	41	130
Rank x clan size	18	-	62
Immigrant	-	41	-
Cohort size	-	-	130
Latency	18	15	89
Approach	18	41	130
Migration	8	17	61
Trial #	6	9	59
Session trial #	6	17	38
Time since last trial	6	16	59
Body condition	2	15	27
Higher rankers present	1	0	38
Trial group size	1	3	10

Note: Greyed cells indicate variables that were "fixed" for inclusion in every model.

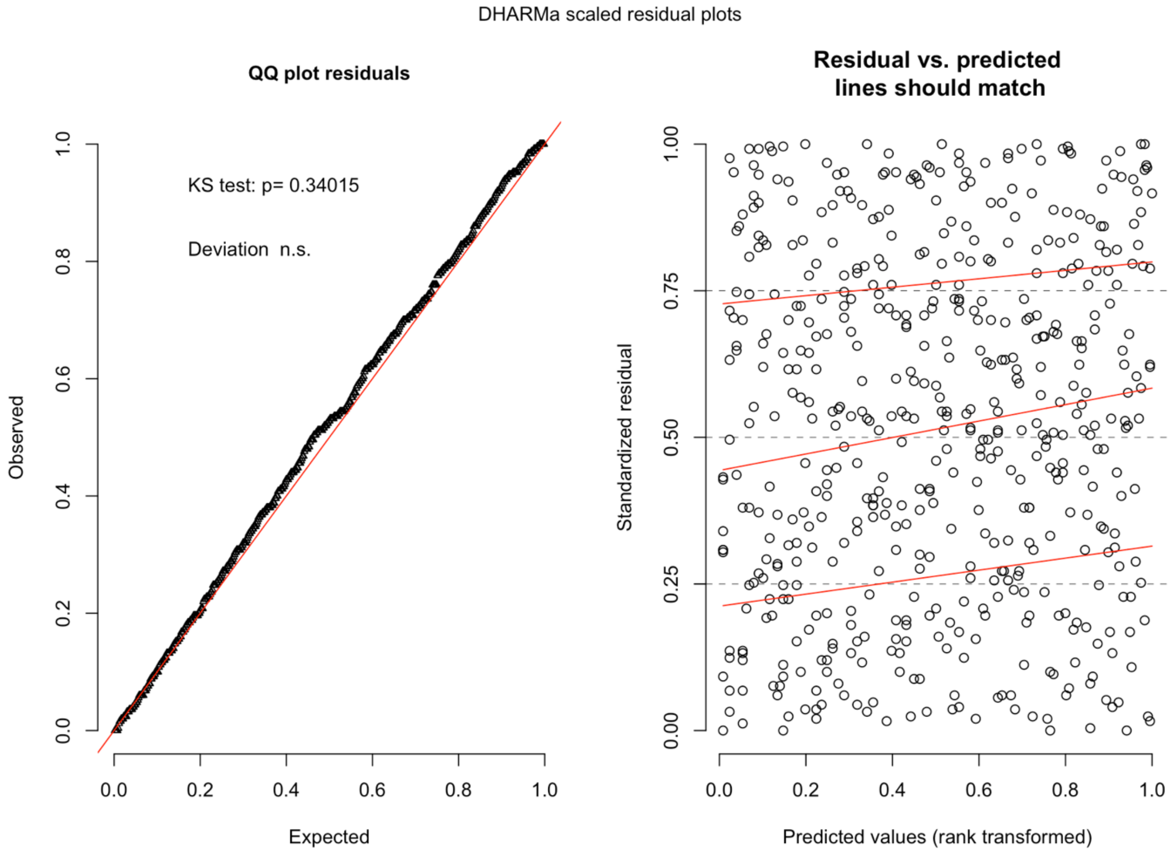

Figure 4.A.1. Residual plots showing qq-plot and a plot of the residuals against the expected values from simulated residuals generated from Model 1 using the R package DHARMa (Hartig, 2019).

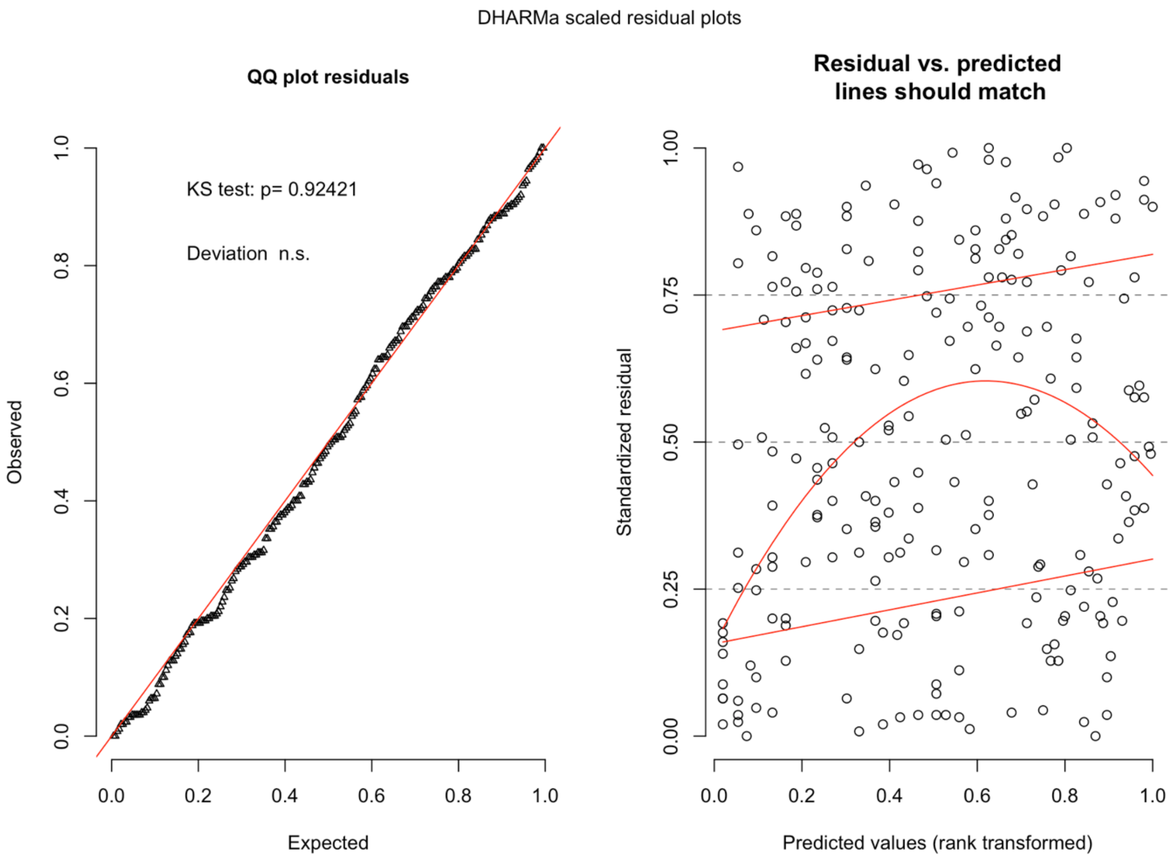

Figure 4.A.2. Residual plots showing qq-plot and a plot of the residuals against the expected values from simulated residuals generated from Model 2 using the R package DHARMa (Hartig, 2019).

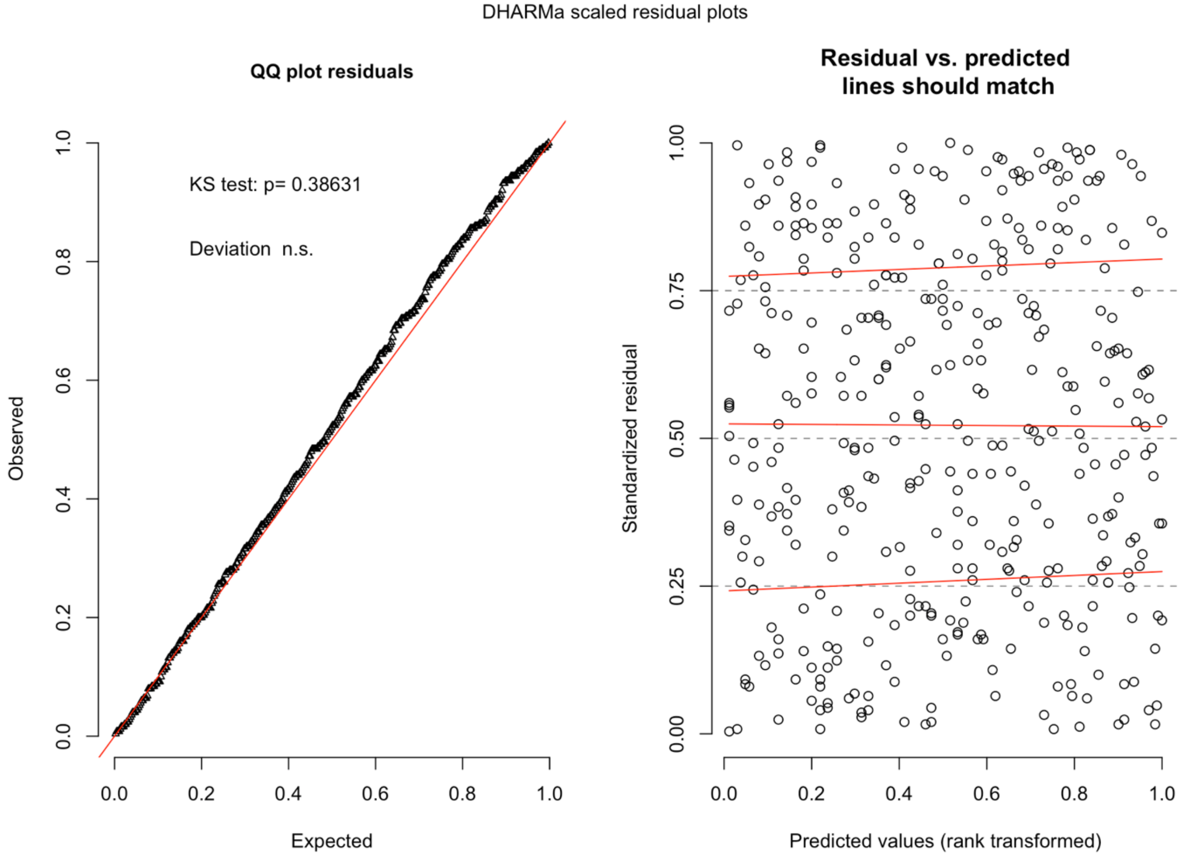

Figure 4.A.3. Residual plots showing qq-plot and a plot of the residuals against the expected values from simulated residuals generated from Model 3 using the R package DHARMa (Hartig, 2019).

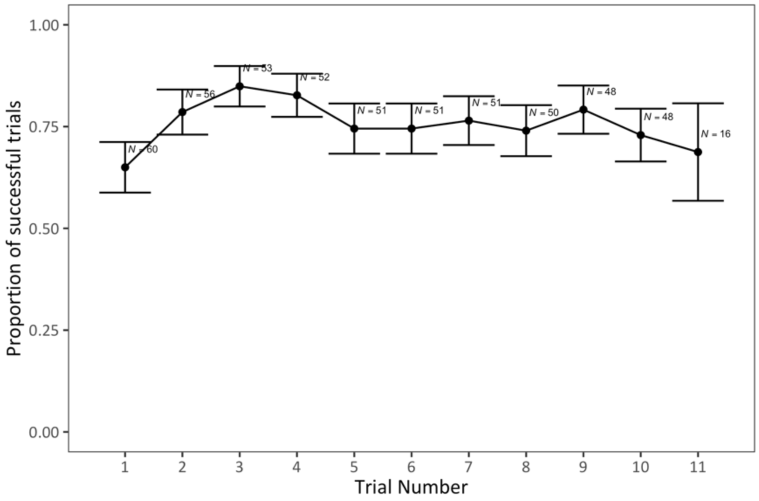

Figure 4.A.4. The relationship between overall trial number with the clear cylinder and proportion of successful trials. Error bars should standard error. Sample sizes (N) indicate the number of subjects that received trials.

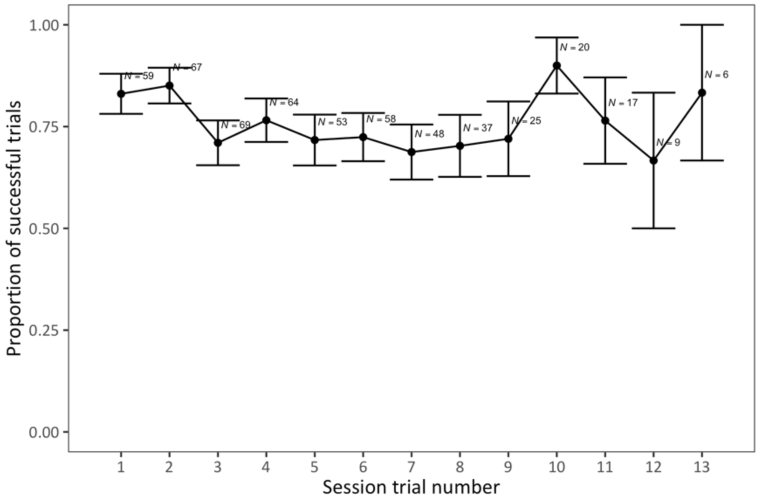

Figure 4.A.5. The relationship between trial number within a testing session and proportion of successful trials with the clear cylinder. Error bars should standard error. Sample sizes (N) indicate the number of subjects that received trials. Session trial number accounts for any trials done with the white cylinder prior to testing with the clear cylinder.

CHAPTER FIVE

URBANIZATION AND INNOVATION IN SPOTTED HYENAS

Lily Johnson-Ulrich, Robyn L. Strong, Gidey Yirga, & Kay E. Holekamp

ABSTRACT

How and why did large brains evolve? The cognitive buffer hypothesis (CBH) suggests that large brains confer a “cognitive buffer” against changing or novel environmental conditions. Support for the CBH comes from research showing that larger brained and more innovative lineages are more likely to successfully invade novel environments, including urban ones. However, it is unclear whether these species were pre-adapted to such environments or whether these environments actually select for increased innovation. Here, we tested innovation in three populations of spotted hyenas experiencing varying degrees of urbanization to test the prediction that increasing urbanization should favor greater innovation because of the high degree of change and novelty the increasing urbanization creates. Spotted hyenas are highly innovative relative to other carnivores and have been established as an excellent model organism for testing the evolution of intelligence. Although many large carnivores are endangered, spotted hyenas remain abundant throughout sub-Saharan Africa. We tested innovation with a multi-access box in which food could be obtained through four different solutions; subjects were scored on the number of solutions learned. Contrary to our predictions, hyenas experiencing increasing urbanization learned fewer solutions, indicating reduced innovation. These results challenged the CBH and support the idea that the novelty or variability created by increasing urbanization does not favor greater innovation.

INTRODUCTION

Innovation is the ability to solve novel problems or use novel behaviors to solve familiar problems (Reader et al., 2016). Innovation represents a crucial aspect of human culture, technology, and creativity (Reader et al., 2016). Although innovation has been documented in a diverse array of species (Reader et al., 2016), it is currently not clear which selective forces may have favored its evolution. The cognitive buffer hypothesis (CBH) is one of the most popular hypotheses put forward to explain the evolution of innovation (Sol, 2009b); the CBH proposes that environmental novelty or variability may favor the evolution of large brains and innovativeness because innovativeness, by allowing animals to develop novel, flexible responses, can buffer the negative effects of novel or variable environments (Sol, 2009b). Support for the CBH comes from research linking brain size and innovation in birds, primates, and carnivores (Benson-Amram et al., 2016; Ducatez, Clavel, & Lefebvre, 2014; Lefebvre, Whittle, Lascaris, & Finkelstein, 1997; Reader & Laland, 2002; Sol et al., 2008), and from research showing that species with larger brains or higher innovation rates are the likeliest to successfully invade novel or climatically variable environments in birds, mammals, amphibians, and reptiles (Amiel et al., 2011; Fristoe & Botero, 2019; Fristoe et al., 2017; Sayol et al., 2016; Schuck-Paim, Alonso, & Ottoni, 2008; Sol et al., 2008, 2005; Sol & Lefebvre, 2000; Sol et al., 2002; van Woerden, van Schaik, & Isler, 2010; van Woerden, Willems, van Schaik, & Isler, 2012; Weisbecker et al., 2015). Although this research supports the idea that innovation is adaptive for responding to evolutionary novelty, recent phylogenetic analyses suggest that the evolution of larger brains in bird predates the invasion of novel and variable habitats (Fristoe et al., 2017; Sayol et al., 2016). These finding suggest that the direction of causality predicted by the CBH might be reversed, and that large brains facilitate the invasion or establishment of bird species in variable or novel habitats. However, this conclusion leaves unanswered the question of which selection pressures may have favored the evolution of large brains and innovative abilities in the original habitats of these animals. Here, we propose to address this question using the intraspecific approach, which allows researchers to identify aspects of the environment that might be cognitively demanding by correlating intraspecific variation in cognition with environmental variation (e.g. Ashton, Thornton, & Ridley, 2018). To do so, we tested innovation in three different populations of spotted hyenas *(Crocuta crocuta)* experiencing different levels of novel environmental conditions caused by urbanization.

Urbanization involves dramatic changes in natural habitats, and presents a host of novel challenges for urban wildlife (Griffin, Netto, & Peneaux, 2017; Sih, 2013; Sol, Lapiedra, & González-Lagos, 2013). Urban environments may differ from natural ones in myriad ways and animals must react to both novel threats and novel opportunities (Griffin, Netto, & Peneaux, 2017). To survive in an increasingly urbanized world, animals must adjust to these novel challenges or face extinction (Sol et al., 2013). The CBH predicts that innovative abilities should be particularly adaptive in urban environments where animals must adjust to novel environmental pressures and research suggests that species with larger brains or overall innovation rates might be more successful at invading urban environments than those with smaller brains (Griffin & Diquelou, 2015; Kark, Iwaniuk, Schalimtzek, & Banker, 2007; Maklakov, Immler, Gonzalez-Voyer, Rönn, & Kolm, 2011; Møller, 2009; Shultz, B. Bradbury, L. Evans, D. Gregory, & M. Blackburn, 2005; Snell-Rood & Wick, 2013). However, results from intraspecific studies that have compared innovative problem-solving between urban and non-urban populations have been mixed; some studies have shown better innovative problem-solving in urban animals (birds: Audet, Ducatez, & Lefebvre, 2016; Griffin et al., 2017;

Preiszner et al., 2017) while some have had ambiguous results or shown no difference between urban and non-urban populations (birds: Griffin et al., 2017; Papp, Vincze, Preiszner, Liker, & Bókony, 2014; Prasher, Evans, Thompson, & Morand-Ferron, 2019; lizards: Kang, Goulet, & Chapple, 2017). However, most of these studies only compared two populations of conspecifics (urban vs rural), and research suggests that selection on brain size or innovation may depend on the stage of invasion, such that selection may be strongest during the initial invasion and colonization, but subsequently decrease after a species becomes established in a novel environment (Griffin et al., 2017; Snell-Rood & Wick, 2013; Wright, Eberhard, Hobson, Avery, & Russello, 2010).

We tested innovative abilities in three different populations of spotted hyenas (*Crocuta crocuta*) experiencing different stages of urbanization. To our knowledge, this is the first study examining the effect of urbanization on innovation in any mammalian species. Spotted hyenas are the most abundant large carnivores in sub-Saharan Africa, where they inhabit almost every ecosystem, including highly urban ones (Baynes-Rock, 2015; Holekamp & Dloniak, 2010; Yirga, Ersino, et al., 2013). Spotted hyenas have sophisticated social cognition and are relatively innovative compared to other hyenids and members of carnivore species (Benson-Amram et al., 2016; Holekamp et al., 2015). In addition, previous research has shown a high degree of intraspecific variation in innovativeness (Benson-Amram & Holekamp, 2012; Johnson-Ulrich et al., 2018), which makes spotted hyenas ideal for examining which environments might favor higher or lower innovativeness. Our first population of spotted hyenas (Population 1) consisted of spotted hyenas living near the center of a protected area, the Maasai Mara National Reserve (MMNR) in Kenya. In this area of the park human activity in the form of tourism, cattle grazing, and poaching is extremely limited. Thus, Population 1 inhabits a relatively stable and pristine environment with little novelty. The environmental conditions under which Population 1 lives are believed to have changed very little since the late Pleistocene (Steele, 2013). Population 2 was also located in the MMNR, but its territory extends beyond the boundary of the park into the rapidly growing Maasai town of Talek. Over the last three decades this population has experienced increasing exposure to cattle grazing, poaching, and tourism (Boydston, Kapheim, Watts, Szykman, & Holekamp, 2003; Green et al., 2018). Telemetry data indicated that, in addition to foraging inside the MMNR, hyenas in Population 2 were indeed using Talek town and surrounding areas to forage on trash and livestock (Green & Holekamp, 2019). Thus, Population 2 represented a transitional habitat that was shifting rapidly from rural to urban (Green et al., 2018), akin to a population in the early stages of colonization of or adaptation to a novel habitat. Population 3 consisted of hyenas living within the city limits of Mekelle, Ethiopia. These hyenas den in some of the only remaining undisturbed forest near Mekelle, in church yards on the edges of the city, and commute daily across agricultural land into the city to forage on trash and livestock remains; most people in northern Ethiopia believe that spotted hyenas eat evil spirits, so the hyenas are not persecuted as they are in Kenya (Yirga Abay et al., 2010; Yirga, Ersino, et al., 2013). Spotted hyenas have inhabited cities in Ethiopia for hundreds of years (Gade, 2006) so Population 3 represents a relatively stable urban environment.

Hyenas in both Populations 1 and 3 have had many generations to adjust to their environments and these environments should pose much less novel or unpredictable challenges for hyenas than do the challenges posed by the rapid change being experienced by hyenas in Population 2. Therefore, the CBH predicts that hyenas in Population 2 should show much higher levels of innovation than those in Populations 1 and 3. Alternatively, because these hyenas are dietary and habitat generalists (Holekamp & Dloniak, 2010) that are relatively innovative compared to other carnivores (Holekamp et al., 2015), it is also possible that we may find reverse causality from what the CBH predicts. That is, it may be that the abilities of spotted hyenas have facilitated their adaptation to urban environments and that the extent of urbanization will not favor greater or lesser innovativeness. This last possibility predicts that all three populations should show equal levels of innovation.

Mechanism of variation across populations

Variation among our populations of spotted hyenas could be caused by micro-evolution, selective immigration of better suited individuals, behavioral plasticity, or all three of these things. Hyenas have inhabited the cities in Ethiopia for over 500 years (Gade, 2006), so there has been ample time for microevolution to have occurred, however, we doubt there has been ample time for microevolution to explain any variation between our populations in the MMNR. Selective immigration as a mechanism of variation is also less plausible for the current study; female spotted hyenas are philopatric whereas male hyenas usually disperse to new clans shortly after reaching sexual maturity; males usually join neighboring clans defending territories that abut the male's natal territory, such that dispersal over large spatial scales is rare (Boydston et al., 2005; Smale, Nunes, & Holekamp, 1997). However, hyenas do show a high level of behavioral plasticity when adapting to anthropogenic disturbance (Boydston et al., 2003; Greenberg & Holekamp, 2017; Julie W Turner, Lafleur, Richardson, & Holekamp, 2019), and we expect that any variation among our three populations is almost certainly caused by behavioral plasticity during development or in response to the immediate context. Ultimately, natural selection may only act on innovative behavior to the extent that socio-ecological conditions favor the expression of innovative behavior (Burkart et al., 2017; Kummer & Goodall, 1985). Thus, explanations of the proximate mechanism or development causes of innovative behavior are not mutually exclusive and only serve to enhance our understanding of which environments favor innovation *and* why they might do so. Thus, our predictions about the role of urbanization on the behavior of hyenas do not vary with the specific mechanism causing variation in innovativeness.

METHODS

Study subjects

We tested 88 free-ranging hyenas across our three populations. All individuals were identified by their unique spot patterns, scarring, and ear damage. Of the 88 tested individuals, 34 were adults and 54 were subadult, 44 were female, 33 were male, and 8 were of unknown sex. Hyenas in the MMNR populations have been observed as part of a long-term research project since 1988, such that we have detailed information about rank, kinship, and demographic variables for every subject. Dates of birth were estimated by the appearance of the natal coat when cubs first appeared above ground and all cubs were sexed based on the morphology of the erect phallus. In Ethiopia, we were able to sex individuals both by the morphology of the erect phallus, when visible, and in adult females by signs of sexual maturity and reproductive activity, such as engorged nipples or a torn phallus. In Ethiopia, we distinguished between two age classes, adults and subadults, based on body size, morphology, pelage length, and signs of sexual maturity in females. Because we could only assign hyenas to one of these two age classes in Ethiopia, we also binned all MMNR hyenas into adult and subadult age classes; we classified subadults as hyenas who were less than three years old and adults as hyenas who were more than three years or had already started breeding. In the MMNR, we tested subjects from two neighboring clans in both Population 1 and Population 2 for a total of four clans in Kenya. In Ethiopia, clan membership was unknown, but testing was conducted at only two sites that were less than a kilometer apart. Subject sampling heavily overlapped at both sites which suggested to us that all hyenas belonged to a single clan. However, it is possible that clan structure is different in urban hyenas than in natural populations (Schramme, 2015).

Test apparatus

We tested innovation using a multi-access box (MAB). Like other innovative problem-solving paradigms or 'puzzle boxes', a MAB tests innovation in the context of extractive foraging (Auersperg et al., 2011). Unlike single-access puzzle boxes, however, MABs typically have multiple entry points, each of which requires unique cognitive skills (Auersperg et al., 2011), unique motor behaviors (Johnson-Ulrich et al., 2018; Manrique et al., 2013) or sequential learning (Huebner & Fichtel, 2015). Our MAB was previously used to test innovation in captive spotted hyenas (Johnson-Ulrich et al., 2018). This MAB was a square galvanized steel box 40.64 x 40.64 x 40.64 cm (length x width x height), weighing approximately 36 kg. Each vertical side of the MAB provided access to the interior of the box via a door that required a different motor behavior to open. Thus, we were able to measure each subject's ability to innovate up to four different times with the four doors, which were as follows: 1) the push door: this was a door 30.5 x 28 cm with a hinge on the side that could be pushed inwards to open. 2) the sliding door: this was a door 30.5 x 28 cm with protruding flanges that could be pushed or pulled sideways to slide open. 3) The pull door: this was a door 30.5 x 28 cm with a hinge on the bottom that could be pulled outwards and downwards to open by grasping a doorknob near the top of the door. 4) The drawer: this was a drawer 10 cm in height that was flush against the bottom of the MAB and took up the entirety of the floor of the MAB; it could be pulled outwards to open. Magnets were used to create mild resistance on all doors to ensure that they would not accidentally fall open; subjects were required to actively interact with the MAB to retrieve the food. All doors could be accessed using either snout or paws. The top of the MAB was removable for familiarization trials. The MAB had multiple 2.5 cm circular holes cut on every side except the bottom so that subjects could smell the food inside during trials. All four doors could be blocked by bolting them shut such that blocked doors could still be manipulated by subjects but would not open to allow food retrieval.

Test procedure

In the MMNR subjects were tested during daily observation periods from 0600 and 1030 or 1700 to 1830 hours – the daylight hours at which hyenas were most active in the MMNR. In Ethiopia, the hyenas were largely nocturnal, therefore, all testing was done between 1800 and 2300 hours – the hours at which hyenas were seen leaving their daytime resting sites and traveling around the city to forage. Research vehicles were used as mobile blinds to hide researchers when deploying the MAB. The MAB was presented opportunistically to hyenas whenever they were found in suitable testing conditions. The MAB was deployed approximately 20m away from subjects and the research vehicle was then moved to a distance of approximately 20m from the MAB. The MAB was deployed preferentially to hyenas in groups of 2 to 5 subjects, but there was some variability in group sizes because hyenas could freely come or go during presentation of the MAB. The average number of hyenas present within 20 m over the entire duration of a trial ± standard deviation was 5.96 ± 4.58 (median = 4, range = 1 – 33). However, the average number of hyenas that actually approached and contacted the MAB over the duration of a trial was 3.54 ± 3.33 (median = 3, range = 1 – 29). Average trial duration was 13.5 ± 15.7 minutes (median = 7 minutes, range = 1 – 100 minutes). In the MMNR, MAB testing locations were categorized as either "den" or "carcass" if either of these were present within 200 m; if neither were present, the location was categorized as "other". In Ethiopia, instead of searching clan territory for hyenas, the MAB was deployed at two sites located along a well-known “commuting” path into the central part of the city of Mekelle. All sessions in Ethiopia were therefore considered “other”.

All trials were videotaped. Trials were initiated when a hyena came within 5 m of the MAB. A trial was considered to have ended if a hyena retrieved the bait, moved at least 5 m away from the MAB for more than 5 minutes, moved at least 100 m away from the MAB, or laid down for 5 minutes without standing up and without contacting the MAB. Hyenas were only considered to have participated in a trial if they made contact with the MAB, and only participating hyenas were given a score for success (opening a door) or failure (failing to open a door) for a trial. Because not participating in a trial did not indicate poor innovation, hyenas who did not contact the MAB during a trial were not given a success or failure score and, instead, they were scored as having not participated. Trial-level data from subjects who did not participate were not included in our analysis of innovation.

If the bait was retrieved and consumed, the MAB was rebaited for consecutive trials. On average, hyenas were given 1.58 ± 1.33 trials per session (median= 1, range = 1 – 11). Consecutive trials were administered as long as hyenas continued consuming the bait or until testing conditions became unsuitable due to external factors such as exhaustion of our bait supply, rain, or darkness.

Prior to testing, hyenas were given familiarization trials in which the MAB was baited but its top was removed and left open. We aimed to give every hyena two familiarization trials prior to being given the option to participate in test trials. On average, hyenas participated in 1.62 ± 1.53 (mean ± standard deviation) familiarization trials prior to being presented with their first test trial, but only fed from the MAB on 0.96 ± 1.11 familiarization trials prior to being presented with their first test trial. However, not all hyenas participated in their first test trials and some hyenas received more familiarization trials after failing to participate in their first test trials. On average, MMNR hyenas fed from the box 2.31 ± 2.15 times during familiarization trials prior to the first test trial in which they participated.

We conducted testing in four different phases for each hyena. During the first phase all four doors were accessible to the subject. Once a hyena used a door in three out of four consecutive test trials, the hyena was considered to have reached criterion for completing that phase of testing and the hyena progressed to phase two. In phase two the door learned in phase one was bolted shut and the subject was required to use one of the three other doors still available to retrieve the bait. In phase three, the doors learned in phases one and two were bolted shut. In phase four, the doors learned in phases one, two, and three were bolted shut.

Subjects were given the opportunity to participate in trials until they either learned all four doors by completing all four phases or they scored five consecutive failures. The MAB was baited with approximately 200 g of either goat or beef muscle, skin, or organ meat. During some trials full cream milk powder was also used in addition to, or in place of, meat. Preliminary analysis of the effect of the type of bait suggests that hyenas were more likely to approach and contact the MAB when it was baited with milk powder than with other types of bait (Table 4.A.1), but the type of bait used had no significant effects on the likelihood of a hyena solving the MAB on any given trial (Table 5.A.1).

Measuring innovation & learning

We measured the performance of hyenas in two different ways. First, each hyena was assigned a score based on the highest phase of testing it achieved. Hyenas received a score of '0' if they failed phase 1 by failing five consecutive trials. Hyenas received a score of '1' if they only opened the box once during Phase 1. A subject received a score of ‘2’ if it reached criterion for phase 1 and progressed to phase 2, a score of ‘3’ if it progressed to phase 3, a score of ‘4’ if it progressed to phase 4, and a max score of ‘5’ if it reached criterion for completing phase 4.

Hyenas that participated in fewer than five test trials, none of which were successful, were not included in our analysis and were not counted among our 88 subjects. In theory, a hyena only had to innovate once in order to demonstrate innovativeness but failing to innovate once is not necessarily good evidence for a lack of innovativeness because a hyena might fail to innovate as a result of various socio-ecological distractions present during trials with wild subjects. We previously analyzed the reliability of innovation using the same dataset in the present analysis (Chapter Three). We found that hyenas were highly reliable in their performance across their first three trials (R = 0.88), but that reliability was lowest when measured across a hyena’s first four trials (R = 0.41) before gradually increasing again (Figure 3.8). Therefore, we erred on the more conservative side and used a cut-off of at least five unsuccessful consecutive trials for inclusion in our analysis. Of these five trials, at least one was required to be a solo test trial with no other hyenas within 20m. In addition, these five trials were conducted across at least two testing sessions. It is possible that these strict inclusion criteria could bias our results if one of our populations had higher levels of participation. For example, if hyenas were more participatory in one of our populations, this could bias the average level of innovativeness downward because more hyenas might reach our criterion of five failures, which would add more scores of zero to the dataset for that population. However, we found no significant differences among populations in participation (Table 3.A.1).

Next, in order to examine variation in the rate at which hyenas reached criterion for each of the four phases of testing and to examine variation in the rate at which they progressed through the phases of testing (i.e. learning of each innovation), we also assigned each hyena a "phase score". Phase scores were calculated as the ratio of the number of successful trials that counted towards criterion to the total number of trials required to pass each phase. This phase score excludes successful trials with a door other than the door a hyena ultimately reached criterion with from the numerator (the door it used in three out of four consecutive trials in that phase). For example, if a hyena had the following sequence of trials: "failure", "failure", "success- drawer", "success-sliding door", "success-sliding door, "success-sliding door", the hyena would receive a phase score of 3 successes with the sliding door out of a total of six trials for a score of 0.5. Hyenas were assigned NAs for phases to which they did not progress. Therefore, a high phase score indicated that a hyena very efficiently reached criterion for that phase whereas a low phase score indicated that hyena was less efficient at reaching criterion for that phase.

Out of 88 hyenas tested, only ten hyenas were included in the dataset with incomplete data; these hyenas failed to reach criterion for ending testing before the end of the study period. However, all ten of these hyenas had solved the MAB on at least one occasion. Because our dataset was zero-inflated (the average innovation score for our 88 subjects was only 0.81 ± 1.58) removing their data would bias each population average downwards to a much greater extent than would including their partial scores.

Statistical analysis

We used Poisson generalized linear mixed models (GLMMs) with the package glmmTMB (Brooks, Mollie et al., 2017) to compare innovation scores among our three populations in the statistical program R (R Core Team, 2019). We created two models; in both the dependent variable was innovation score. In the first model we included population identity and age class as independent variables. In our second model, we restricted the dataset only to hyenas from Populations 1 and 2 so we could examine the potential influence of additional independent variables. In addition to population identity and age class, this model included sex, social rank, clan size, an interaction between age class and sex, and an interaction between rank and clan size. We included these two interactions because adult females in a clan are all natal animals whereas most adult males in a hyena clan are immigrants and this status might potentially have an effect on innovative propensity. Next, we included an interaction between rank and clan size because previous research found that low ranking animals in larger clans had better inhibitory control, which is an ability thought to be related to innovation (Chapter Four; Johnson-Ulrich & Holekamp, 2019). Model fit was assessed using the R package Dharma (Hartig, 2019).

Next, we used latent growth curve modeling to examine variation, and the correlates of any such variation, in subject performance across phases in the software program Mplus (Version 8). Latent growth curve modeling is a type of structural equation model that uses latent variables, called growth factors, to model the change or development of an outcome over time. We used latent growth curve modeling in addition to our initial analysis because compressing each hyena's performance across phases to just a single innovation score ignores the variation in subject performance across phases, even among subjects with the same final score. Using latent growth modeling allowed to us capture the variation in growth that led to each subject's final score and examine subtler differences between our three populations of hyenas. Here, instead of assigning each hyena a single score, we scored hyenas based on performance across each of the four phases of testing. We wanted to compare hyenas who learned more quickly and took fewer trials to reach the criterion for progressing to the subsequent phase than did those hyenas who required a larger number of trials to move forward. Each of the four phases of testing were therefore used as our four time points for modeling growth. In latent growth curve modeling, the shape of the average growth curve across the entire data set is estimated by selecting the best fitting model without any covariates. The intercept (I) and slope (S) (growth factors) of this model can then be compared to covariates of interest and used to assess how well they explain variation in the data set. For more complex models, additional growth factors, such as quadratic terms, may also be calculated. The best fitting model is identified by comparing several fit indices in tandem, the χ^2 goodness of fit test, adjusted Bayesian information criterion (ABIC), the comparative fit index (CFI), the Tucker-Lewis index (TLI), root mean square error of approximation scores (RMSEA) and standardized root mean square residual scores (SRMR). The χ^2 test compares the observed data to the estimated model's data. Here, a significant p-value indicated a significantly large discrepancy between the observed and estimated data and thus indicated poor fit. Therefore, a non-significant p-value does not indicate good fit per se, but instead it rules out significantly poor fit. Lower ABIC scores, as with Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) indicate better fit relative compared to other models within the planned set of models. ABIC scores penalize models for adding more parameters. The CFI uses χ^2 scores and adjusts for the sample size to resolve the issues of large sample sizes in the χ^2 test. A CFI value of 1 would indicate perfect fit and a 0 would indicate poor fit. The TLI is similar in conception to the CFI, and scores are interpreted in the same manner. The SRMR test compares the estimated model to a saturated model, or a model in which every possible parameter is estimated. For SRMR a score of less than 0.05 is desired.

The ultimate goal of our analysis was to quantify any differences, or lack thereof, among our three populations of hyenas that experienced various degrees of urbanization. Urbanization was treated as an ordinal variable indicating degree of urbanization with a score of 1 for Population 1, 2 for Population 2, and 3 for Population 3. Once we'd identified the best fitting model with our time points, we regressed the growth factors (e.g. intercept and slope) onto the degree of urbanization. That is, degree of urbanization was included in a regression model as the independent variable with the intercept and slope as dependent variables. These growth factors were thus used to indicate innovative performance in our populations. For example, a positive estimate for urbanization on slope would indicate that hyenas in Population 3 had a larger slope (faster rate of growth) than those in the other populations. A positive estimate for urbanization on the intercept would indicate a higher performance in phase 1 by hyenas in Population 3. We also systematically identified other possible categorical covariates including, clan membership, age class, and sex by including them in regression models in isolation and finally with each other and with degree of urbanization. Once again, we chose the best fitting model by comparing several metrics in tandem. Ultimately, based on our predictions we expected that degree of urbanization would have a positive effect on our growth factors.

RESULTS

We tested 88 wild hyenas across our three populations, 35 in Population 1 (rural), 38 in Population 2 (transitional), and 15 in Population 3 (urban). On average hyenas earned an innovation score of 0.81 ± 1.58 (median = 0, range = 0 – 5). Contrary to our expectations we found that hyenas in Population 2 (GLMM: $z = -3.13$, $P = 0.002$) and Population 3 (GLMM: $z = -3.26$, $P = 0.001$) both had significantly lower innovation scores than hyenas in Population 1 (Figure 5.1; Table 5.1). In addition, age class also had a significant effect, where subadults generally had lower innovation scores than adults (GLMM: $z = -2.58$, $P = 0.01$).

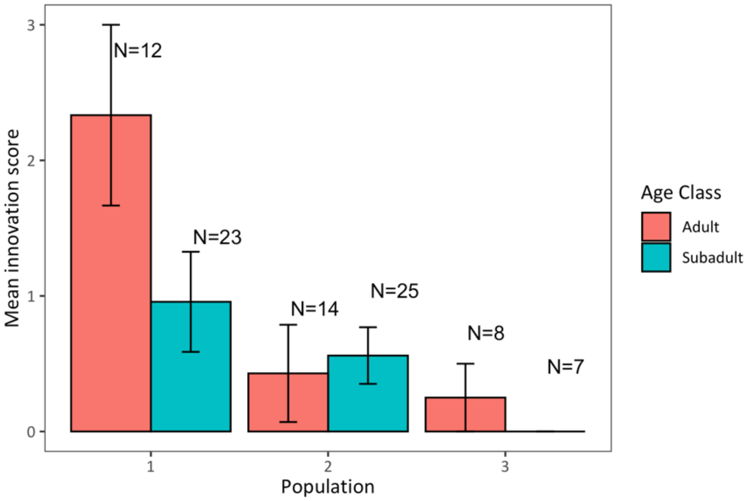

Figure 5.1. Average innovation scores for each population and age class. Error bars show standard error.

Table 5.1. Results from a GLMM on innovation scores in populations 1, 2, and 3.

	β	**SE**	***z***	***P***
Intercept	0.7135	0.1858	3.841	<0.001
Population 2	-1.035	0.2646	-3.911	<0.001
Population 3	-2.4887	0.7225	-3.445	<0.001
Age class – subadult	-0.61	0.2362	-2.583	0.010

In our next model, we only compared the scores of hyenas from populations 1 and 2 (our rural and transitional populations from the MMNR) in order to examine the effect of other factors on innovation. However, none of our additional independent factors were significant (Table 5.2). Once again, Population 2 (transitional) had significantly lower innovation scores than Population 1 (rural) (GLMM: $z = -2.94$, $P = 0.003$) and subadults had significantly lower innovation scores than adults (GLMM: $z = -2.76$, $P = 0.006$).

Table 5.2. Results from a GLMM comparing innovation scores in Populations 1 and 2.

	β	**SE**	*z*	***P***
Intercept	-0.58	1.26	-0.46	0.65
Population 2	-0.95	0.32	-2.94	0.003*
Age class – subadult	-1.10	0.40	-2.76	0.006*
Sex – male	-0.35	0.53	-0.66	0.51
Rank	0.73	0.59	1.24	0.22
Clan size	0.01	0.01	0.94	0.35
Age class * Sex	1.10	0.67	1.63	0.10
Rank * Clan size	-0.01	0.01	-1.14	0.25

*Indicates statistical significance at α = 0.05

Latent growth curve analysis

The mean phase score across all populations and all phases was 0.11 (SD = 0.27) although the average score varied considerably within each phase (Figure 5.2).

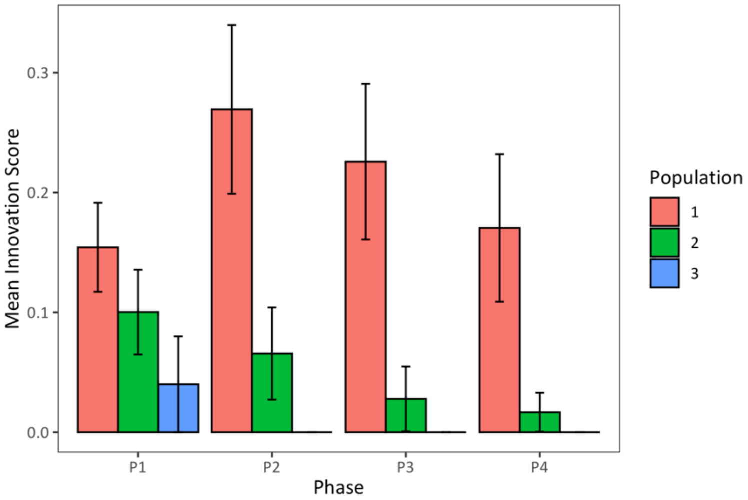

Figure 5.2. Mean innovation scores for each population across the four phases of testing; error bars show standard error. A score of 1 indicates high innovation and a score of 0 indicates low innovation.

We ran a series of models to determine the overall shape of the growth curve across phases prior to including any covariates. We first began with an intercept- and slope-only model (2 growth factors) before adding a quadratic term (3 growth factors). We also ran linear and quadratic models with freed time scores, where the loadings for the slope growth factor at each time score are allowed to vary (nonlinear and non-quadratic models). These models had four dependent variables: phases one through four, and two or three latent variables: the growth factors. Overall, a nonlinear model provided the best overall fit (Table 5.A.2). Next, we added covariates to the nonlinear model to see if any of them improved model fit on their own. If multiple covariates improved model fit, we also included them in another model together to examine their relative influence. These covariates included: degree of urbanization (indicated by population number), clan membership with populations, age class, and sex. We examined model fit by examining several fit indices in conjunction: the χ^2 test, ABIC, CFI, TLI, and SRMR scores.

Degree of urbanization, age class, and clan member ship dramatically improved model fit when added as covariates, and they also had significant effects on the intercept and slope. When sex was added to the nonlinear model, the model was unable to converge, likely because sex was unknown for most subjects in Population 3. Overall, we selected a nonlinear model with degree of urbanization and age class as our final model. Adding clan to this model did not significantly improve model fit, and in this model clan was not significant, which suggests that degree of urbanization and age class explained more residual variance that clan could alone. The model without clan was also more parsimonious because fewer parameters were estimated.

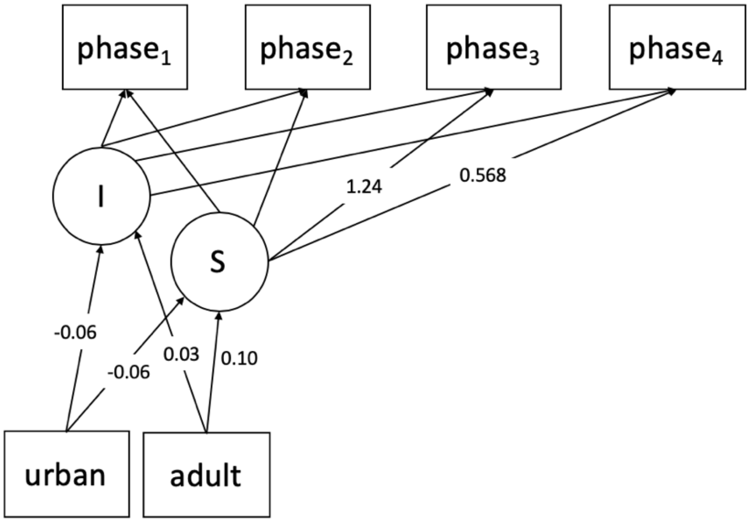

Figure 5.3. Final best-fitting model with path coefficients for estimated parameters above paths. Growth was modeled across four time points: phases 1 through 4. Variables inside circles are the latent growth factors: intercept (I) and slope (S). The intercept was fixed at 1 for all for phases. The slope was fixed at zero and one for phases one and two but was freely estimated at phases three and four (nonlinear growth). Arrows between urban or adult and I or S show the estimates for the effect of each on the growth factors.

Overall, our final model had good fit (Table 5.A.2). The χ^2 value was not significant, but this model had the lowest ABIC score, CFI and TLI values were both close to one and higher than other models, and the SRMR was less than 0.05. In this model, there were 4 dependent variables (phases one through four), two independent variables (degree of urbanization and age class), and two continuous latent variables (intercept and slope) (Figure 5.3). This model had 14 free parameters; the intercept for phases one through four were all fixed at one, the slope was fixed at zero and one for phases one and two but was freely estimated at phases three and four. Residual variance was freely estimated for three dependent variables and both latent variables; the variance for phase three was fixed at zero because it was extremely close to zero. Both the slope and the intercept of this model were positive and significant (Table 5.3) which indicated that, on average, there was significant growth across the four phases of testing.

Table 5.3. Final latent growth curve model results.

	β	***P***
Mean I	0.206	<0.001*
Mean S	0.095	0.03*
S WITH I	0.014	0.02*
I ON urban	-0.063	0.04*
I ON age	0.100	0.46
S ON urban	-0.062	0.01*
S ON age	0.100	0.01*

*Indicates statistical significance at α = 0.05

In addition, the correlation between the slope and the intercept was positive, indicating that hyenas who started with higher baseline innovation scores in phase one, had larger growth across phases than those with lower baseline innovation scores. Degree of urbanization had a significant negative effect on both the intercept and slope growth factors which indicated that hyenas experiencing a higher degree of urbanization started off with lower scores on phase one and also underwent less growth across phases. Effect of age class was positive for both the slope and intercept, but only significant for slope which suggests that adult and subadult hyenas did not perform differently on phase one, but that adult hyenas showed greater improvement across phases than subadults.

DISCUSSION

Our results did not support the CBH's prediction that innovation scores would be highest in Population 2, which was the population currently undergoing the transition from rural to urban. Instead, we found that average innovation scores were significantly higher in the rural

Population than the transitional population, and we found that transitional and urban populations did not differ significantly from one another. We also found that degree of urbanization had a significant negative impact on initial performance in phase one and on growth across phases.

Overall, these results challenge the CBH's assertion that novel environments, such as those created by urbanization, are cognitively demanding. Instead, our results initially appear to support the reverse causality hypothesis which suggests that large brains and innovativeness predate the invasion of variable environments. Spotted hyenas are relatively innovative compared to other carnivores and this innovativeness may indeed be linked to their ability to colonize so many habitats in sub-Saharan Africa, including urban ones. This conclusion resembles that of phylogenetic studies of brain size in birds, which found that large brains and innovative abilities correlated with successful colonization of novel habitats, but also that the evolution of large brains predated the invasion of these novel environments (Fristoe et al., 2017; Sayol et al., 2016). However, neither of our two hypotheses predicted that hyenas in Population 1 would have significantly higher innovativeness. Just as it is unclear what environmental factors favored the evolution of large brains in bird lineages in their original habitats, it also remains unclear what kinds of environmental factors favor higher levels of innovative ability in spotted hyenas in our rural population.

Physiological correlates of intraspecific variation in innovation

First, it is possible that urbanization correlates with reduced innovation in spotted hyenas because a physiological mechanism, such as stress or hunger, is playing a role. Previous research with spotted hyenas in the MMNR has found higher stress levels, measured using fecal glucocorticoids, in populations experiencing more anthropogenic disturbance (Van Meter et al., 2009), and we cannot rule out the possibility that stress caused the reduced innovative abilities in our transitional population here. However, our urban population has been established in northern Ethiopia for many generations, so it would be surprising not to find an attenuated stress response to anthropogenic activity in this population (Atwell et al., 2012; Partecke, Schwabl, & Gwinner, 2006). Hunger can also play an important role in the motivation to obtain food and it possible that trash and livestock remains presents a more abundant food resource for hyenas in our urban population in Ethiopia relative to hyenas in the transitional or rural populations in the MMNR. For hyenas, body condition generally indicates how recently a hyaena has eaten (Kruuk, 1972). However, in the present study body condition was not a significant predictor of trial success; 'fat' hyenas were not less likely to have a successful trial with the MAB (Binomial GLMM: $z = -1.39$, $P = 0.16$; Table 5.A.1). Furthermore, when we created a Poisson GLMM with body condition as a numeric response variable with scores ranging from 0-2 and population identity as a categorical predictor variable, we found no significant differences in body condition among the three populations (Poisson GLMM: $z = 0.27$, $P = 0.79$). We also attempted to control for factors sch as motivation, or neophobia, that could affect participation in trials with the multi-access box by only including subjects in our analyses if they had at least one successful trial or at least five consecutive unsuccessful trials. Our criterion for scoring a trial as unsuccessful required a hyena to fully participate in a trial by making contact with the MAB, which should have excluded hyenas that weren't properly motivated. Finally, research that has investigated how either limited or abundant resources can promote innovation has found only mixed support for the idea that necessity, created by resource limitations, can promote innovation (Griffin & Guez, 2014; Reader et al., 2016). On the other hand, the hypothesis that abundant resources, by facilitating spare energy or time, can promote innovation has received much stronger support (Benson-Amram et al., 2013; Kummer & Goodall, 1985; Reader et al., 2016). This 'spare time' hypothesis would have predicted hyenas in our urban population to express higher rates of innovation, rather than lower.

Environmental correlates of intraspecific variation in innovation

Our results suggest that how animals respond to urbanization is likely to be related to their adaptive match with urban environments (Griffin et al., 2017; Sol, 2008) and the role that these environments have on the development and acquisition of cognitive skills . Although cities are evolutionarily novel to all animals, including humans, they might not actually pose any novel challenges for spotted hyenas, and may in fact reduce the demand for innovation. For hyenas, urbanization may pose more opportunities than challenges per se because scavenging from trash dumps is likely far less challenging than hunting wild prey. Hyenas in both Kenyan populations hunt and kill at least 90% of the prey in their diet (Holekamp et al., 1997), whereas “hunting” in the form of livestock depredation accounts for only 11% to 15% of total diet of hyenas in urban Ethiopia (Yirga Abay et al., 2010). Within mammalian carnivores, diet and hunting behavior were correlated with increases in overall brain size (Swanson, Holekamp, Lundrigan, Arsznov, & Sakai, 2012), so it is possible that the reduced demand for hunting in our transitional and urban populations due to the presence of anthropogenic food sources is associated with the development of reduced innovativeness in our study populations. This hypothesis would be supported if rates of hunting or scavenging correlated with innovativeness within or between populations. Second, a composite measure of sociality that included group size, subgroup size, and social cohesion, was correlated with relative brain size within carnivora (Swanson et al., 2012). Although hyena density (Holekamp & Dloniak, 2010; Yirga et al., 2017) appears to increase with increasing urbanization, social structure might also change with urbanization (Belton, Cameron, & Dalerum, 2018), and preliminary research suggests there is reduced social cohesion among hyenas in our transitional and urban populations (dissertations: Schramme, 2015; Julie Wolfe Turner, 2018). Innovation appears to correlate with sociality across many animals potentially as a result of the benefits of social learning or social facilitation (Ashton, Ridley, et al., 2018; Ashton, Thornton, & Ridley, 2019; Dunbar & Shultz, 2017; Griffin & Guez, 2015; Muthukrishna, Doebeli, Chudek, & Henrich, 2018; van Schaik et al., 2012). This 'cultural' hypothesis would be supported if greater social cohesion in our rural population facilitated the acquisition of innovative behavior in a greater number of individuals and led to a higher rate of innovation in this population. Third, lion density also covaries with urbanization across our three populations. Spotted hyenas in our rural population coexist with a higher density of lions than those in our transitional population (Green et al., 2018), and lions are locally extinct in the range of cities of northern Ethiopia (Yirga, De Iongh, et al., 2013). Historically, lions represent the main source of mortality for spotted hyenas (Green et al., 2018; Watts & Holekamp, 2009), and some versions of the CBH suggest that large brains may help buffer animals from any source of mortality, whether it is from environmental variability, predation, or other sources (Allman et al., 1993; Sol, Székely, Liker, & Lefebvre, 2007). Thus, it is possible that a high degree of interspecific conflict with lions may drive innovativeness in hyenas in our rural population. We hope to test these hypotheses in the future by correlating data on hunting and scavenging behavior, social networks, lion-hyena conflict, and innovativeness both within and between our study populations. Ultimately, urbanization creates myriad changes in the social and physical landscape and the way in which urbanization alters the environment my vary between Kenya and Ethiopia. In the present study we compared only a single urban, a single transitional, and a single rural population of hyenas and although we have discussed the aspects of the environment that we suspect may be most relevant to hyena behavior and innovation, it is likely that we have not considered all the factors that may vary across our three populations in a way that affects the expression of innovative behaviors. Furthermore, because of this limited sample size at the population level it is very possible that our results will not generalize to other hyena populations. For example, fully urbanized hyenas in Ethiopia, where their presence is relatively well-tolerated, might express a very different response to urbanization than hyenas in other African countries. Thus, not only should researchers carefully consider the ecologically relevant aspects of urbanization to their study species, but future work should aim to compare the same species across multiple urban or transitional populations.

Conclusion

In sum, our finding that increased urbanization was associated with decreased innovation warrants further investigation, not only within spotted hyenas, but also in other animals experiencing various degrees of urbanization. This research was the first work to examine the effects of urbanization on innovative problem-solving in a mammal; to the best of our knowledge, previous research has been exclusively on birds (Federspiel et al., 2017; Kark et al., 2007; Papp et al., 2014; Prasher et al., 2019; Preiszner et al., 2017). Thus, our work represents an important step towards determining how urbanization might affect mammals and larger-bodied species. Indeed, the body size of animals may dramatically influence how they respond to urbanization and the scale at which urbanization affects them (Abelson, 2016; Bateman & Fleming, 2012). Our research was also the first to compare innovativeness among populations experiencing three different levels of urbanization in the same species; our finding that hyenas living in both transitional and fully urban environments had lower innovation scores challenges an assumption of the CBH that novel environments favor innovation, and calls for more careful quantification of differences between urban and non-urban habitats. Finally, because urbanization is having an increasing impact on animals around the globally, further research is of utmost importance into understanding how or why some animals are able to adapt to it, while others are rapidly going extinct.

APPENDIX

Table 5.A.1. Binomial GLMM model results on the factors affecting the probability of solving the MAB. Dependent variable was a binomial factor indicating whether a hyena solved or failed to solve the MAB.

	β	**SE**	***z***	***P***
(Intercept)	-16.53	4.42	-3.74	<0.001*
Sex – male	1.33	1.94	0.69	0.49
Age class - cub	1.75	1.94	0.90	0.37
Age class – adult	2.47	1.40	1.77	0.08
Rank	0.55	0.92	0.59	0.55
Trial #	0.56	0.16	3.56	<0.001*
Phase #	1.90	0.95	2.01	0.04*
# of times previously fed from the MAB	0.43	0.39	1.11	0.27
Body condition – fat	-1.26	0.90	-1.39	0.16
Body condition - skinny	-15.56	5798.62	-0.00	0.99
Bait - rotten	1.368	0.83	1.64	0.10
# hyenas present w/in 20m	0.07	0.07	0.93	0.35
# hyenas present that contacted the MAB	-0.25	0.12	-2.15	0.03
Bait – milk	0.34	1.00	0.34	0.74
Bait – muscle tissue	0.02	0.89	0.03	0.98
Bait – bone	-1.14	0.64	-1.80	0.07
Bait – offal	-1.47	1.04	-1.42	0.16
Trial # x Phase #	-0.22	0.06	-3.71	<0.001*

*Indicates statistical significance at $\alpha = 0.05$.

Table 5.A.2. Model fit indices for all latent growth curve models fitted.

Model	**# free parameters**	**χ^2 value**	**χ^2 p-value**	**ABIC**	**CFI**	**TLI**	**SRMR**
Linear	9	49.912	0.00	-177.51	0.84	0.81	0.14
Nonlinear	10	13.05	0.01	-181.496	0.97	0.95	0.06
Quadratic	9	44.78	0.00	-182.64	0.86	0.83	0.13
Non-quadratic	11	38.77	0.00	-186.01	0.88	0.75	0.14
Nonlinear + Urban	12	13.90	0.03	-218.76	0.97	0.96	0.06
Nonlinear + Age	12	13.6	0.03	-218.34	0.97	0.96	0.05
Nonlinear + Sex	13	NA	NA	NA	NA	NA	NA
Nonlinear + Clan	12	15.21	0.02	-217.48	0.97	0.95	0.06
Nonlinear + Urban + Age	**14**	**14.57**	**0.07**	**-226.11**	**0.98**	**0.96**	**0.05**
Nonlinear + Urban + Age + Clan	16	17.41	0.07	-225.27	0.98	0.96	0.04

NOTE: Final selected model indicated in bold.

CHAPTER SIX

FITNESS CONSEQUENCES OF INNOVATION IN SPOTTED HYENAS

Johnson-Ulrich, L., Benson-Amram, S., & Holekamp, K. E. (2019). Fitness consequences of innovation in spotted hyenas. *Frontiers in Ecology and Evolution*, *7*, 443. https://doi.org/10.3389/fevo.2019.00443

ABSTRACT

Innovation is a well-studied cognitive phenomenon related to general intelligence and brain size. Innovative ability varies considerably within species and it is widely assumed that this variation must have important fitness consequences. However, direct evidence for a link between innovative ability and fitness has rarely been shown. Previous research examined variation in innovative problem-solving in wild spotted hyenas when confronting a novel puzzle box baited with meat. The earlier work revealed that variation in innovativeness in spotted hyenas was not related to age, sex, or social rank, but was predicted by neophobia, persistence, and diversity of motor responses to the puzzle. Here, we used the same dataset from wild spotted hyenas to investigate potential links between innovativeness and fitness. We found that innovative hyenas had lower offspring survivorship than non-innovators, but higher annual cub production (ACP). To test the hypothesis that high ACP can compensate for low offspring survival, we also measured annual cub survivorship (ACS) counting only offspring that survived at least 1 year. Here, there was no significant difference between innovators and non-innovators, which suggests that higher ACP does compensate for lower offspring survival, at least to 1 year of age. Overall, our data suggest that innovativeness may have both costs and benefits for fitness in wild spotted hyenas.

INTRODUCTION

Innovation, solving a novel problem, or using a novel behavior to solve a familiar problem, is widely studied in humans and animals (Kummer & Goodall, 1985; Reader & Laland, 2003). Innovation has been strongly linked to brain size across bird, primate, and carnivore species (Benson-Amram et al., 2016; Lefebvre et al., 2004; Reader & Laland, 2002). Innovation is also thought to be an important marker of high general intelligence (Ramsey, Bastian, & van Schaik, 2007; Reader et al., 2016) across a diverse array of taxa. Whereas the socio-ecological causes of inter- and intraspecific variation in innovative ability have been well-studied (reviewed in Reader & Laland, 2003; Reader et al., 2016), the fitness consequences of variation in innovativeness have rarely been examined despite growing interest (Ashton, Ridley, et al., 2018; Boogert et al., 2018; Morand-Ferron et al., 2015).

Innovative ability is typically measured on the species level by observing the rate of spontaneous innovations demonstrated in the wild (Lefebvre, Reader, & Sol, 2013), and on the individual level by experimentally presenting captive or wild subjects with novel problem-solving tasks (Griffin & Guez, 2014; Reader et al., 2016), which typically require performance of a novel behavior to obtain a reward. Research suggests that innovativeness may be beneficial for adjusting to novelty and environmental change (Sol, Sayol, Ducatez, & Lefebvre, 2016), an idea supported by correlations between innovation rates and generalist dietary or habitat preferences (Ducatez et al., 2014; S. E. Overington, Griffin, Sol, & Lefebvre, 2011). Furthermore, species that are more innovative appear more likely to invade novel habitats, including urban ones (Griffin & Diquelou, 2015; Sol et al., 2005). Finally, more innovative bird species tend to have slower life-histories and longer lifespans (Sol, 2009a; Sol et al., 2016). Overall, this work suggests that innovativeness is likely adaptive for individuals responding to environmental change and novelty by enabling those individuals to express novel behaviors, exploit novel food sources, or avoid novel sources of mortality. Likewise, innovativeness is generally assumed to increase fitness through enhanced survival or reproductive success by buffering individuals against mortality-causing events (Sol, 2009a; Sol et al., 2016), by increasing mating success (Chen, Zou, Sun, & ten Cate, 2019; Keagy, Savard, & Borgia, 2009), or by increasing foraging rate, efficiency or quality. However, direct evidence supporting these assumptions is scarce.

Although innovation and general intelligence in humans have been consistently related to positive life outcomes (Plomin & Deary, 2015), the relationship between innovative problem-solving and fitness in non-human animals is much less clear. Across bird species, five studies have found a positive link between innovative problem-solving and fitness measures that included mating success, clutch size, hatching success, fledgling survival, provisioning rates, and offspring survival (Cauchard, Boogert, Lefebvre, Dubois, & Doligez, 2013; Chen et al., 2019; Keagy et al., 2009; Preiszner et al., 2017; Wetzel, 2017). However, other studies of birds found no relationship between innovative problem-solving and mating success (Isden, Panayi, Dingle, & Madden, 2013), or found that innovative problem-solving was correlated with lower competitiveness and higher nest desertion (Cole, Morand-Ferron, Hinks, & Quinn, 2012; Cole, Quinn, Cole, & Quinn, 2012). In the only study that has looked at innovation and fitness in a mammal, Huebner, Fichtel, and Kappeler (2018) found no link between more efficient problem-solving and any measure of fitness in mouse lemurs. Overall, the literature linking innovation and fitness in animals is very small, with limited taxonomic representation, and with largely mixed results. Here our goal was to examine the relationship between innovativeness and fitness in wild spotted hyenas.

Spotted hyenas are large African carnivores that have previously been established as a good model system for testing hypotheses about the evolution of cognition (Holekamp et al., 2007). Unlike most large carnivores in Africa, spotted hyenas are not endangered; their success may have been facilitated by their impressive behavioral flexibility. Spotted hyenas are generalist feeders; they eat everything from termites to elephants (Cooper, Holekamp, & Smale, 1999; Hayward, 2006) and have established themselves in nearly every habitat in sub-Saharan Africa (Holekamp & Dloniak, 2010) including urban ones (Yirga et al., 2017). Earlier research found that spotted hyenas show innovative ability similar to that of wild vervet monkeys (Benson-Amram & Holekamp, 2012), and that spotted hyenas also show high levels of innovativeness relative to other carnivores (Benson-Amram et al., 2016). In the present study we aimed to test the idea that innovativeness might be an adaptive trait in spotted hyenas by comparing their problem-solving performance to three measures of fitness. To do this, we used a subset of the data from Benson-Amram & Holekamp (2012) on innovative problem-solving in female spotted hyenas, and analyzed it in relation to our long-term data on reproduction and survival for each female. This subset included 29 female spotted hyenas, of which five were considered innovative.

Hyena fitness has been linked to both social and ecological variables. Dominance rank has large effects on lifetime reproductive success in hyenas; the highest ranking female in a clan may have up to five times more offspring than the lowest ranking female due to better access to food, younger ages at first parturition, shorter interbirth intervals, better offspring survival, and longer reproductive lifespans (Frank, 1986; Holekamp et al., 1996). Finally, ecological variables such as prey abundance and competition with lions also affect reproductive success and juvenile survival (Watts & Holekamp, 2009). Our goal here was to test the hypothesis that innovativeness is adaptive in regard to both reproductive success and survival in wild spotted hyenas; if true, we expected to see a direct positive relationship between innovative problem-solving and our measures of fitness.

METHODS

Subjects, population, and location

The subjects were individuals from two neighboring clans (the Talek West clan and the Fig Tree clan) of spotted hyenas in the Maasai Mara National Reserve, Kenya. Individuals were identified by their unique spot patterns and other natural markings. Observations were conducted daily from 0530 to 0900 h and from 1700 h to 2000 h, on an average of 23.5 days per month. The Talek West clan was monitored continuously from May 1988 to December 2016, and the Fig Tree clan was monitored continuously from April 2007 to May 2015. All innovation testing took place between May 2007 and May 2008; during this period, the Talek West clan contained 46–48 members, including 12–13 adult females with their juvenile offspring and 10 adult males, and the Fig Tree clan contained 36–38 members, including 10 adult females with their juvenile offspring and 7–8 adult immigrant males. Additional information about the study subjects, methods and materials can be found in Benson-Amram & Holekamp (2012). Although innovativeness was tested in both male and female hyenas, in the current analysis we only included female hyenas for which we had reproductive data.

Problem-solving apparatus

We used a novel problem-solving apparatus to test innovative ability. The “puzzle box” used here measured 60 × 31 × 37 cm and was built from welded 10.5 mm steel rebar (Figure 6.1). The box had a single 30 × 34 cm door on one long side, large enough for a hyena to put its head inside the box and handles in the center of each short side. When it was baited with roughly

2 kg of raw meat, the box weighed more than 35 kg. To obtain access to the meat, a subject had to slide a 12 cm steel bolt latch laterally using the mouth or forepaws, and the door would swing open. For more detail on the apparatus, see Benson-Amram and Holekamp (2012).

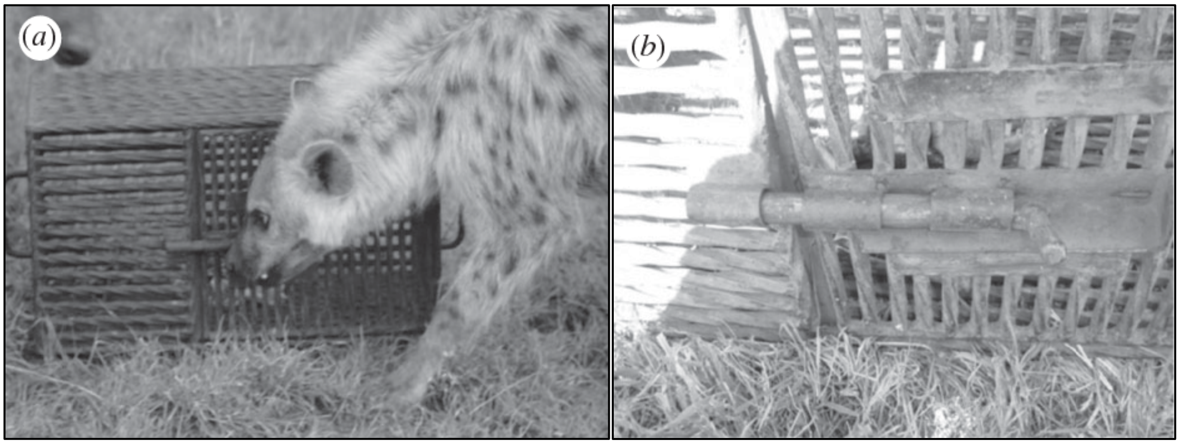

Figure 6.1. A) An image of the puzzle box apparatus used in the experiment. **B)** A close-up image of the latch bolt that hyenas had to move laterally in order to access the meat inside the puzzle box. Republished from (Benson-Amram & Holekamp, 2012).

Successful trials were those in which the puzzle box was opened. Unsuccessful trials included those in which the hyena contacted the box, but failed to open it, as well as those in which the hyena did not interact with the box, despite spending time within 5 m of it (average duration in minutes spent within 5 m on the first trial ± SD = 11.95 ± 13.47, N = 29). Previously, we found that 14.5% of all hyenas tested with this problem-solving task had at least one successful trial. Within this group of successful hyenas, 78% were successful on subsequent trials. Trial number was a significant predictor of latency to solve the problem, with hyenas generally solving the box faster in later trials, which suggests that hyenas learned how to open the box (see Figure 2 in Benson-Amram and Holekamp, 2012). Successful problem-solving also showed modest but significant repeatability after controlling for the effect of trial number (Likelihood ratio test: R = 0.24, $P < 0.0001$) (rptR package; (Stoffel et al., 2017). Additionally, in Chapter Three, with the same population of wild spotted hyenas, we found that innovation was significantly repeatable across four novel problem-solving tasks (Likelihood ratio test: R = 0.97, $P < 0.001$; Table 3.1: Model 1.1). Therefore, in the present study hyenas were defined as innovative if they had at least one successful trial and non-innovative if they had only unsuccessful trials.

Data collection protocols

Because we were working with a wild population, subjects for these experiments were chosen opportunistically, based on which animals were available at the time. However, every attempt was made to conduct equal numbers of trials with all the individuals in each clan, and to balance the number of participants in each age, sex and social rank category. When an appropriate subject animal was sighted in an accessible location, we parked our research vehicle ~100 m upwind of the hyena. The box was placed on the ground on the opposite side of the vehicle from the hyena in a location with good visual access, both for the subject, and for observers. The box was oriented with the door toward the hyena, with the latch protruding at 90° from the box, parallel to the ground. We then pulled the vehicle back ~50 m from the box and initiated observations. A trial began when a hyena approached to within 5 m of the box (thereby becoming a "focal hyena"); the trial ended when the hyena left the 5 m radius and remained outside of it for 5 min, or when it moved to at least 200 m from the box. All attempts were made to test subjects only when they were alone, but occasionally conspecifics approached and participated in a trial. However, the presence or absence of conspecifics did not significantly affect the likelihood of a hyena successfully opening the box (Benson-Amram and Holekamp, 2012), so this variable was not analyzed in the present study. All trials were videotaped in their entirety from our vehicle. For more detail on data collection protocols see Benson-Amram and Holekamp (2012).

Fitness variables

Demography

We used several demographic variables to calculate survivorship and annual reproductive success. First, cub ages were estimated to ± 1 week based on their appearance when first observed (Holekamp et al., 1996), and date of birth (DOB) was calculated on that basis. Date of first conception (DFC) for each adult female was estimated by subtracting 110 days, the length of gestation in spotted hyenas, from the DOB of a female's first observed litter. Some females in the Fig Tree clan were adults when monitoring began; therefore, DFC represents the first conception that researchers observed, but might not represent every female's first conception after reaching sexual maturity. Finally, date last seen (DLS) was recorded as the last day on which a hyena was seen alive or, if its body was found, the date on which it was found dead. Female hyenas remain in their natal clans throughout their lives (Kruuk, 1972) so females that had not been seen for at least 6 months were considered to be dead. Standardized social rank was measured as a continuous variable on a scale of 1 to −1 where a rank of 1 indicated the highest-ranking female in the clan and a rank of −1 indicated the lowest-ranking female in the clan. All individuals in a clan were assigned their own rank except for pre-weaning cubs and subadults who were assigned the rank of their mother. Ranks were assigned based on the clan hierarchy during the period from May 2007–2008, when innovation was tested. This hierarchy was generated using a dominance matrix ordering observations of aggressive or submissive behaviors within dyads of adult hyenas (Holekamp et al., 2012; Martin & Bateson, 1993). Rank hierarchies among spotted hyenas are convention-based such that offspring acquire ranks immediately below those of their mothers through a process of maternal interventions and social learning. Social ranks in spotted hyenas are relatively stable and rank reversals are rare (Strauss & Holekamp, 2019).

Offspring survivorship

Offspring survivorship was calculated from birthdate and mortality data during the first 24 months of life. Mortality was recorded as a binary variable: dead vs. alive. Using this mortality data, the proportion of offspring surviving was estimated at each age (in months) up to 24 months of age. If offspring disappear before 24 months of age, this represents mortality, but this is not necessarily true after 24 months because 24 months represents the age at which hyenas reach sexual maturity and male hyenas begin to disperse then from their natal clans (Van Horn et al., 2003). Thus, disappearance after 24 months of age for male hyenas may be due to either mortality or dispersal.

Annual reproductive success

Offspring survivorship does not necessarily correlate with lifetime reproductive success because it does not account for the number of offspring produced. Therefore, the next measure of fitness we examined was annual reproductive success. We included two measures of annual reproductive success: annual cub production (ACP) and annual cub survival (ACS). ACP was calculated by dividing the total number of cubs born to a female during the study divided by her observed reproductive lifespan. Observed reproductive lifespans were calculated by subtracting a female's DFC from her DLS or the end date of the study. Annual cub survival (ACS) was calculated in the same manner as ACP, but instead of using the number of cubs born, only the number of cubs surviving to 1 year of age were counted.

Statistical analysis

All statistical analyses were done using R version 3.5.0 (R Core Team, 2019). To analyze offspring survival we used a Cox proportional hazards model, which is ideal for analyzing right-censored time-to-event data. This model estimated the probably that subjects would survive to specific ages by using both the lifespan and mortality variables described in section Offspring Survivorship. Cox regression was conducted using the R packages “survival” (Therneau, 2015; Therneau & Grambsch, 2000) and “survminer” (Kassambara & Kosinski, 2018). For all other fitness analyses we used linear regression models built using the R package “glmmTMB” (Brooks, Mollie et al., 2017). The dependent variables in each of our models were survival, ACP, or ACS. Every model included innovativeness as the independent variable. Subject rank, number of trials, and an interaction effect between innovativeness and number of trials were included as potential confounds in all models. We included rank to control for its previously demonstrated effect on reproductive success in spotted hyenas (Holekamp et al., 1996). We included the number of trials each hyena received prior to her first successful trial in each model to control for the number of opportunities each hyena had to open the puzzle box. If the hyena had no successful trials, this number represented the total number of trials in which she participated. Likewise, we added an interaction effect because subjects who solved the puzzle box on their first trial were potentially demonstrating a higher level of innovative ability than those who solved the box after many trials. That is, the effect of innovativeness on fitness might depend on the trial number. We also included a random effect of maternal ID in the Cox regression. Because proportions such as ACP and ACS might not fully account for the potentially confounding effect of length of the observed reproductive lifespan, we also included the length of the observed reproductive lifespan as a covariate in these two models. Full output from each model is available in the Appendix (Tables 6.A.1–6.A.3). Model fit for each model was assessed using the R package "DHARMa" (Hartig, 2019). All models showed good fits as indicated by normally distributed residuals, non-significant DHARMa non-parametric dispersion tests, and non-significant Durbin-Watson tests for temporal autocorrelation.

RESULTS

Innovative problem-solving

Thirty-three female hyenas participated in trials with the puzzle box; however, the exact social ranks of two females were unknown, and two other females had incomplete reproductive data, so they were dropped from the analysis, yielding a sample size of 29 (Table 6.1). Of these 29 females, five females were able to open the box at least once and were thus considered to be innovative. On average, female hyenas received 4.48 ± 4.16 trials (range = 1 – 14 trials) and opened the box an average of 1.62 ± 4.53 times (range = 0 – 18).

Table 6.1. Summary of number of subjects in each rank and age class combination.

	High-ranking	**Mid-ranking**	**Low-Ranking**	**Totals**
Pre-weaning subadult	3	3	1	7
Post-weaning subadult	3	1	0	4
Adult	9	7	2	18
Totals	15	11	3	29

Offspring survivorship

These 29 females produced 288 offspring across the study period, and we confirmed 114 cases of mortality within the first 24 months of age. Using a Cox proportional hazards regression model, we found that offspring of innovative mothers had significantly lower survival rates during the first 24 months than offspring of non-innovative mothers (Cox: $z = 2.31$, $P < 0.02$; Figure 6.2; Table 6.A.1). None of the other covariates were significant in this model.

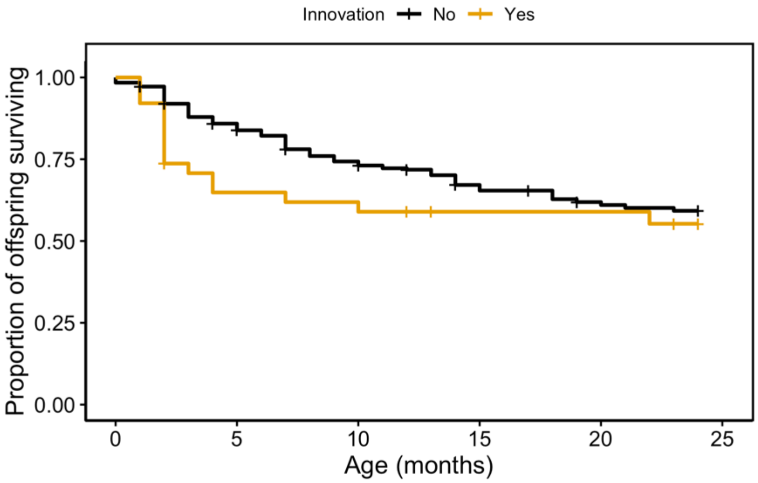

Figure 6.2. Unadjusted survival curves calculated showing the proportion of offspring surviving at each age point between 0 and 24 months of age for non-innovative and innovative mothers.

Annual reproductive success

On average, subjects gave birth to 9.6 ± 4.48 cubs (range = 3 – 19) during their observed reproductive lifespan. The average length of the observed reproductive lifespan was 8.24 ± 4.57 years (range = 1.73 – 20.00). Across all 29 females, without controlling for covariates, average ACP was 1.29 ± 0.35 cubs per year (range = 0.60 – 1.93). In our model, innovative females produced significantly more cubs annually than did non-innovative females (LM: $z = 2.85$, $P = 0.004$; Figure 6.3A; Table 6.A.2). Innovative females produced an average of 1.35 ± 0.24 (estimated marginal mean ± SE) cubs annually, whereas non-innovative females produced only 1.2 ± 0.07 cubs annually. The length of the observed reproduction lifespan (LM: $z = -3.42$, $P < 0.001$) and the interaction between innovativeness and trial number were also significant in this model (Table 6.A.2). The effect of innovativeness on ACP was highest for female hyenas with the fewest trials (LM: $z = -0.44$, $P = 0.01$). When this interaction effect was not included the model, the effect of innovation on ACP was not significant (LM: $z = 1.04$, $P = 0.30$).

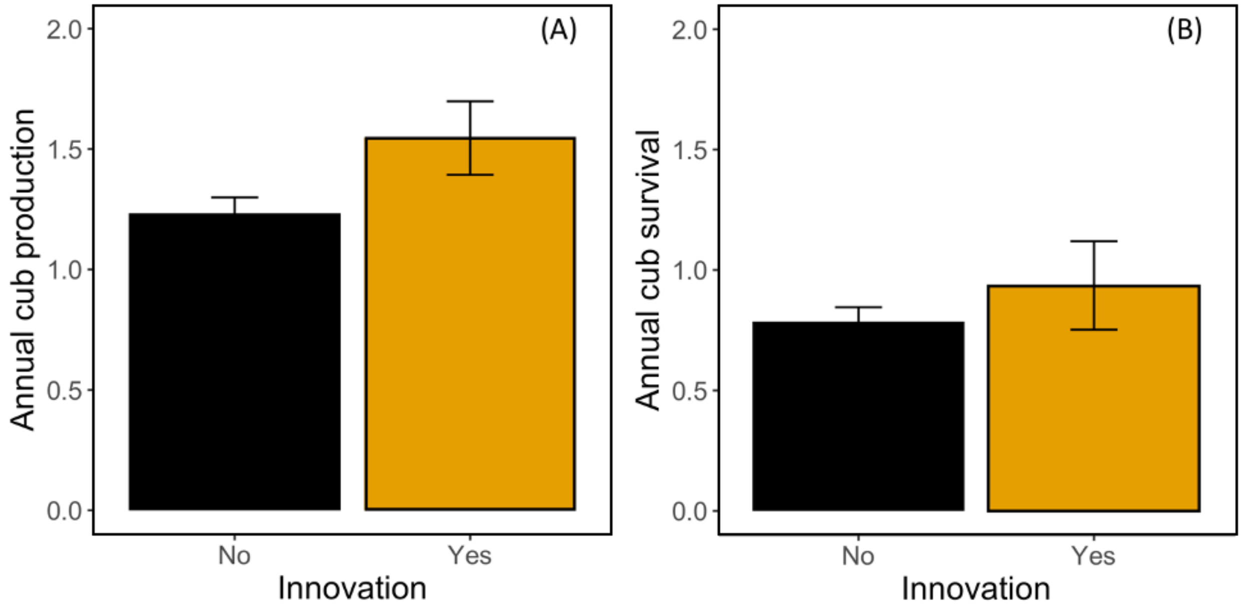

Figure 6.3. Annual reproductive success for innovative and non-innovative females. **A)** Average number of cubs produced annually by female hyenas. **B)** Overall number cubs that survive to 1 year of age for female hyenas. Errors bars show standard error.

Next, we investigated the possibility that high ACP could compensate for low offspring survival by comparing the overall number of cubs surviving to 1 year of age between innovative and non-innovative females. On average, without controlling for covariates, females produced 0.81 ± 0.31 surviving cubs each year (range = 0 – 1.62). In our model, innovative females produced 0.94 ± 0.18 cubs annually and non-innovative females produced 0.79 ± 0.06 cubs annually, but this difference was not significant (LM: $z = 1.28$, $P = 0.20$), suggesting that innovative and non-innovative females produce similar numbers of cubs that survive to 1 year of age (Figure 6.3B; Table 6.A.3). No other covariates were significant in this model (Table 6.A.3).

DISCUSSION

We found that innovativeness was linked to fitness in variable ways in wild spotted hyenas. Innovative females had lower offspring survivorship, but gave birth to more offspring annually, than did non-innovative females. In addition, innovative and non-innovative females gave birth to similar numbers of cubs that survived to 1 year of age. This suggests that innovative females might be able to offset lower offspring survival with higher reproductive success. However, given our very small sample size of 29 hyenas (of which, only five were able to solve the puzzle box), our results should be interpreted cautiously. Our results are similar to those from other species suggesting that cognitive abilities may have both fitness costs and benefits. For example, guppies that were artificially selected for larger brains had better performance on cognitive tasks, but produced fewer offspring (Kotrschal et al., 2013a), and more innovative great tits had larger clutches but also had higher levels of nest desertion (Cole, Morand-Ferron, et al., 2012). Fitness trade-offs involving non-cognitive traits appear to be fairly common in animals (e.g. Barrickman, Bastian, Isler, & van Schaik, 2008; Ducatez, Audet, & Lefebvre, 2019; Lewin, Swanson, Williams, & Holekamp, 2017; Ricklefs & Wikelski, 2002; Sinervo, Svensson, & Comendant, 2000; Wolf, van Doorn, Leimar, & Weissing, 2007), so it should probably not surprise us to find that this may be true in regard to cognitive traits as well.

Life history trade-offs

Because innovative females had lower offspring survival (Figure 6.2), but higher ACP (Figure 6.3A) than non-innovative females, we considered the possibility that these correlations represented an adaptive trade-off between two alternative life history strategies where innovative female hyenas pursue a faster, quantity over quality, life history strategy and non-innovative females pursue a slower, quality over quantity, life history strategy. Previous research in wild spotted hyenas found that in juveniles, high insulin-like growth factor 1 (IGF-1) concentrations correlated with fast growth and earlier reproduction, but shorter lifespans (Lewin et al., 2017), suggesting that hyenas may invest differentially in reproduction and survival. However, innovation is generally thought to be associated with slower life histories across species (Sol et al., 2016). Instead of, or in addition to, alternative life-history strategies, it is also worth considering the possibility that innovativeness is an evolutionary stable strategy where the adaptive benefits of innovativeness are only realized at a specific ratio of innovative to non-innovative hyenas in the study population due to frequency-dependent selection. It is conceivable that the presence of a small ratio of innovative females could be beneficial at the scale of the entire clan if innovative females make previously unexploited resources available to the entire clan. Although spotted hyenas show only limited social learning of novel behaviors, feeding itself is highly socially facilitated (Benson-Amram et al., 2014; Yoerg, 1991).

Our result showed that innovative and non-innovative females have similar numbers of cubs surviving to 1 year, which suggests that higher birth rates in innovative females do indeed offset the significantly lower offspring survival rates. For spotted hyenas, mortality rates are highest in the first year of life; nearly half of all cubs born perish during their first year (Watts & Holekamp, 2009). Therefore, it seems reasonable to assume that the proportion of cubs that survive to 1 year would correlate with the overall proportion of offspring surviving to sexual maturity for female hyenas. This could be interpreted as evidence for equal adaptive value between being innovative vs. non-innovative; however, previous research on spotted hyenas found that, whereas annual reproductive success significantly predicted lifetime fitness, the length of the reproductive lifespan was the strongest determinant of lifetime fitness in spotted hyenas (Swanson, Dworkin, & Holekamp, 2011). We were unable to calculate the actual reproductive lifespans of our female subjects because our dataset was both right and left-censored temporally; some of our subjects were adults females without known DOBs or dates of sexual maturity, and many were still alive at the end of the study period. However, the average observed reproductive lifespan in our dataset was 8.24 ± 4.57 years which, although censored for some subjects, is not significantly different from the average reproductive lifespan in our study population (7.13 ± 3.34 years, N = 170) (Swanson et al., 2011). Therefore, it is possible that innovative and non-innovative females have similar lifetime fitness, but, without actual lifetime fitness data on a larger sample of females, we cannot conclude this with any certainty.

Mediators of the link between innovativeness and fitness

Not only are researchers interested in the links between cognition and fitness, but also the mechanisms mediating such linkages. However, it is often unclear just why a specific cognitive ability might improve reproductive success or survival. The largest natural source of mortality for wild spotted hyenas comes from lions, both directly through conflict and indirectly through competition over food resources (Watts & Holekamp, 2009). In hyenas and other animals innovative problem-solving has been linked to greater boldness or risk-taking behavior (Audet et al., 2016; Benson-Amram & Holekamp, 2012; Johnson-Ulrich et al., 2018; Sarah E Overington, Cauchard, Côté, & Lefebvre, 2011; van Horik et al., 2017; Webster & Lefebvre, 2001), which in turn are correlated with higher mortality in wild hyenas. Hyenas that are bolder in the presence of lions, in particular, have a higher risk of mortality than conspecifics with intermediate or low levels of boldness (Yoshida et al., 2016), so lower survivorship among offspring of innovative females may be mediated by high boldness during conflict or competition with lions. In addition, greater risk-taking behavior in spotted hyenas, measured with a "mock intruder" test, is also correlated with a higher risk of mortality (Julie W Turner et al., 2019). Overall, if more proactive, bold, or risk-taking behavior, demonstrated by hyenas while interacting with problem-solving apparatuses, is correlated with their behavior in other contexts, it is possible that these traits mediate the link between innovation and survival.

The relationship between innovative problem-solving and reproductive success, on the other hand, has been linked in some bird species to the ability to forage more efficiently (Cauchard et al., 2017; Preiszner et al., 2017; Wetzel, 2017) but see (Cole, Morand-Ferron, et al., 2012). Access to food is a strong determinant of reproductive success among female hyenas; social rank is the strongest determinant of reproductive success because high ranking individuals enjoy the best access to high quality food resources (Holekamp et al., 1996). In addition, both average fatness, which usually indicates how recently a hyena has fed, and per capita prey availability also correlate with reproductive success in hyenas (Swanson et al., 2011; Watts & Holekamp, 2009). Social rank does not predict innovativeness in spotted hyenas (Benson-Amram & Holekamp, 2012); therefore, if innovativeness is correlated with the ability to access food in hyenas, then it is plausible that this would directly increase reproductive success.

Assumptions and limitations

Our analysis of the relationship between offspring survival and female innovativeness is based on only a small sample and hinges on several assumptions. First, for innovativeness to be related to offspring survival, innovativeness must be transmitted from mother to offspring through genetic heritability or social learning. However, few studies have assessed the heritability of innovative problem-solving and one that has, in great tits, found no evidence for heritability (Quinn, Cole, Reed, & Morand-Ferron, 2016). If instead the relationship between innovative ability and offspring survival is mediated by a trait such as boldness, then boldness must be transmissible. Although the heritability of innovativeness in wild spotted hyenas has not been tested, previous research indicates that their boldness is heritable (Yoshida et al., 2016). Alternatively, it is also possible that innovative ability is entirely stochastic and has a direct effect on offspring survival through early-life effects or ongoing social support. Cubs usually wean between 12 and 18 months of age, but female hyenas provide ongoing social support to their mature female offspring throughout their lives during feeding competition and other social interactions with clan members (Smith et al., 2010; Vullioud et al., 2019; Watts & Holekamp, 2009). Thus, it is possible that variation in this support, if related to the ability to innovate, results in differential survival between offspring of innovative and non-innovative females.

Although our analysis of annual reproductive success in spotted hyenas does not hinge on assumptions about heritability, it is less robust than our analysis of offspring survival due to an extremely small sample size and censored windows of time during which we were able to monitor reproductive output for many subjects. Our total sample size consisted of 29 female hyenas, of which only five were innovators. Of these five individuals three were high ranking, one was mid-ranking, and one was low-ranking. In addition, three were adults at the time innovativeness was measured and two were pre-weaning subadults. Thus, these five innovative hyenas did not differ in any substantial measurable way from our overall sample of females, but our results should still be interpreted with caution because we cannot be sure that these five are not outliers in ways we did not measure. Maternal age might also have affected our measures of fitness. For individuals with known birth dates, the length of the reproductive lifespan would have controlled for this. Unfortunately, without knowing the birth dates for some of the females sampled (N = 2 innovators, N = 5 non-innovators), we have no way of knowing what their age was during the observed portions of their reproductive lives.

Conclusion

In summary, innovative female spotted hyenas were found to have lower offspring survival, but higher annual cub production, than non-innovative females. These results suggest there might be trade-offs among the costs and benefits of innovativeness, or that innovative and non-innovative females pursue different life history strategies. We would benefit from further study of the relationship between fitness and innovativeness in wild spotted hyenas. Ideally, long-term study would allow for measurement of lifetime reproductive success and assessing the heritability of innovativeness across generations. In addition, future work might investigate the mediators of the relationship between innovativeness and fitness by comparing innovativeness to other variables such as foraging ability, boldness, and social relationship.

APPENDIX

Table 6.A.1. Results from the cox proportional hazards regression.

Dependent variable	Hazard ratio	β	SE	*z*	*P*
Innovation	2.03	0.71	0.45	2.26	0.02*
Trials	1.01	0.01	0.02	0.48	0.63
Rank	1.16	0.14	0.16	1.01	0.32
Innovation x trials	0.70	-0.35	0.46	-1.01	0.31

*Indicates statistical significance at α = 0.05.

Table 6.A.2. Results from the linear regression on annual cub productivity.

Dependent variable	β	SE	*z*	*P*
Innovation	0.61	0.23	2.68	0.007*
Trials	0.00	0.01	0.19	0.85
Innovation x trials	-0.44	0.17	-2.54	0.01*
Rank	0.09	0.08	1.23	0.22
Repro years	-0.04	0.01	-3.42	0.0006*

*Indicates statistical significance at α = 0.05.

Table 6.A.3. Results from the linear regression on annual cub survivorship.

Dependent variable	β	SE	*z*	*P*
Innovation	0.35	0.27	1.28	0.20
Trials	-0.00	0.01	-0.22	0.83
Innovation*trials	-0.16	0.21	-0.74	0.46
Rank	-0.01	0.09	-0.08	0.93
Repro years	0.01	0.01	-1.03	0.30

*Indicates statistical significance at α = 0.05.

REFERENCES

REFERENCES

Abelson, E. S. (2016). Brain size is correlated with endangerment status in mammals. *Proceedings of the Royal Society B: Biological Sciences*, 283(1825), 20152772-. https://doi.org/10.1098/rspb.2015.2772

Allman, J., McLaughlin, T., & Hakeem, A. (1993). Brain weight and life-span in primate species. *Proceedings of the National Academy of Sciences*, 90(1), 118–122. https://doi.org/10.1073/pnas.90.1.118

Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Current Biology*, 18(18), 1415–1419. https://doi.org/10.1016/j.cub.2008.08.020

Amici, F., Call, J., & Aureli, F. (2009). Variation in withholding of information in three monkey species. *Proceedings of the Royal Society B: Biological Sciences*, 276(1671), 3311–3318. https://doi.org/10.1098/rspb.2009.0759

Amici, F., Call, J., Watzek, J., Brosnan, S., & Aureli, F. (2018). Social inhibition and behavioural flexibility when the context changes: a comparison across six primate species. *Scientific Reports*, 8(1), 3067. https://doi.org/10.1038/s41598-018-21496-6

Amiel, J. J., Tingley, R., & Shine, R. (2011). Smart moves: effects of relative brain size on establishment success of invasive amphibians and reptiles. *PloS One*, 6(4), e18277. https://doi.org/10.1371/journal.pone.0018277

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 13(3), 214–228. https://doi.org/10.1177/1073858407299288

Arsznov, B. M., Lundrigan, B. L., Holekamp, K. E., & Sakai, S. T. (2010). Sex and the frontal cortex: A developmental CT study in the spotted hyena. *Brain, Behavior and Evolution*, 76(3–4), 185–197. https://doi.org/10.1159/000321317

Ashton, B. J., Ridley, A. R., Edwards, E. K., & Thornton, A. (2018). Cognitive performance is linked to group size and affects fitness in Australian magpies. *Nature*, 554(7692), 364–367. https://doi.org/10.1038/nature25503

Ashton, B. J., Thornton, A., & Ridley, A. R. (2018). An intraspecific appraisal of the social intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), 20170288. https://doi.org/10.1098/rstb.2017.0288

Ashton, B. J., Thornton, A., & Ridley, A. R. (2019). Larger group sizes facilitate the emergence and spread of innovations in a group-living bird. *Animal Behaviour*, 158, 1–7. https://doi.org/10.1016/j.anbehav.2019.10.004

Atwell, J. W., Cardoso, G. C., Whittaker, D. J., Campbell-Nelson, S., Robertson, K. W., & Ketterson, E. D. (2012). Boldness behavior and stress physiology in a novel urban environment suggest rapid correlated evolutionary adaptation. *Behavioral Ecology*, 23(5), 960–969. https://doi.org/10.1093/beheco/ars059

Audet, J.-N., Ducatez, S., & Lefebvre, L. (2016). The town bird and the country bird: problem solving and immunocompetence vary with urbanization. *Behavioral Ecology*, 27(2), 637–644. https://doi.org/10.1093/beheco/arv201

Auersperg, A. M. I., Gajdon, G. K., & Bayern, A. M. P. Von. (2012). A new approach to comparing problem solving , flexibility and innovation. *Communicative & Integrative Biology*, 5(2), 140–145. https://doi.org/10.4161/cib.18787

Auersperg, A. M. I., von Bayern, A. M. P., Gajdon, G. K., Huber, L., & Kacelnik, A. (2011). Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PloS One*, 6(6), e20231. https://doi.org/10.1371/journal.pone.0020231

Aureli, F., & Schino, G. (2019). Social complexity from within: how individuals experience the structure and organization of their groups. *Behavioral Ecology and Sociobiology*, 73(1), 6. https://doi.org/10.1007/s00265-018-2604-5

Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology*, 108, 44–79. https://doi.org/https://doi.org/10.1016/j.pneurobio.2013.06.005

Barrett, L., Henzi, P., & Dunbar, R. I. M. (2003). Primate cognition: from “what now?” to “what if?” *Trends in Cognitive Sciences*, 7(11), 494–497. https://doi.org/10.1016/j.tics.2003.09.005

Barrett, L., Henzi, P., & Rendall, D. (2007). Social brains, simple minds: does social complexity really require cognitive complexity? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 561–575. https://doi.org/10.1098/rstb.2006.1995

Barrickman, N. L., Bastian, M. L., Isler, K., & van Schaik, C. P. (2008). Life history costs and benefits of encephalization: a comparative test using data from long-term studies of primates in the wild. *Journal of Human Evolution*, 54(5), 568–590. https://doi.org/10.1016/j.jhevol.2007.08.012

Bartoń, K. (2018). MuMIn: Multi-Model Inference. Retrieved from https://cran.r-project.org/package=MuMIn

Barton, R. A., & Dunbar, R. I. M. (1997). Evolution of the social brain. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (Vol. 2, pp. 240–263). Cambridge, United Kingdom: Cambridge University Press.

Bateman, P. W., & Fleming, P. A. (2012). Big city life: carnivores in urban environments. *Journal of Zoology*, 287(1), 1–23. https://doi.org/10.1111/j.1469-7998.2011.00887.x

Baynes-Rock, M. (2015). *Among the Bone Eaters: encounters with hyenas in Harar* (Vol. 8). Penn State Press.

Behaviour, A. (2017). Guidelines for the treatment of animals in behavioural research and teaching. *Animal Behaviour*, 123, I–IX. https://doi.org/10.1016/S0003-3472(16)30351-7

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, 77(4), 771–783. https://doi.org/10.1016/j.anbehav.2008.12.022

Belton, L. E., Cameron, E. Z., & Dalerum, F. (2018). Social networks of spotted hyaenas in areas of contrasting human activity and infrastructure. *Animal Behaviour*, 135, 13–23. https://doi.org/10.1016/j.anbehav.2017.10.027

Benson-Amram, S., Dantzer, B., Stricker, G., Swanson, E. M., & Holekamp, K. E. (2016). Brain size predicts problem-solving ability in mammalian carnivores. *Proceedings of the National Academy of Sciences*, 113(9), 2532–2537. https://doi.org/10.1073/pnas.1505913113

Benson-Amram, S., Heinen, V. K., Gessner, A., Weldele, M. L., & Holekamp, K. E. (2014). Limited social learning of a novel technical problem by spotted hyenas. *Behavioural Processes*, 109, 111–120.

Benson-Amram, S., & Holekamp, K. E. (2012). Innovative problem solving by wild spotted hyenas. *Proceedings of the Royal Society B: Biological Sciences*, 279(1744), 4087–4095. https://doi.org/10.1098/rspb.2012.1450

Benson-Amram, S., Weldele, M. L., & Holekamp, K. E. (2013). A comparison of innovative problem-solving abilities between wild and captive spotted hyaenas, *Crocuta crocuta*. *Animal Behaviour*, 85(2), 349–356. https://doi.org/10.1016/j.anbehav.2012.11.003

Bergman, T. J., & Beehner, J. C. (2015). Measuring social complexity. Animal Behaviour, 103, 203–209.

Bjorklund, D. F., & Harnishfeger, K. K. (1995). 5. The evolution of inhibition mechanisms and their role in human cognition and behavior. In F. Dempster & C. Brainerd (Eds.), *Interference and Inhibition in Cognition* (pp. 141–173). Elsevier. https://doi.org/10.1016/B978-012208930-5/50006-4

Boogert, N. J., Madden, J. R., Morand-Ferron, J., & Thornton, A. (2018). Measuring and understanding individual differences in cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), 20170280. https://doi.org/10.1098/rstb.2017.0280

Borrego, N., & Dowling, B. (2016). Lions (*Panthera leo*) solve, learn, and remember a novel resource acquisition problem. *Animal Cognition*, 19(5), 1019–1025. https://doi.org/10.1007/s10071-016-1009-y

Boucherie, P. H., Loretto, M.-C., Massen, J. J. M., & Bugnyar, T. (2019). What constitutes "social complexity" and "social intelligence" in birds? Lessons from ravens. *Behavioral Ecology and Sociobiology*, 73(1), 12. https://doi.org/10.1007/s00265-018-2607-2

Bousquet, C. A. H., Petit, O., Arrivé, M., Robin, J.-P., & Sueur, C. (2015). Personality tests predict responses to a spatial-learning task in mallards, *Anas platyrhynchos*. *Animal Behaviour*, 110, 145–154.

Boydston, E. E., Kapheim, K. M., Van Horn, R. C., Smale, L., & Holekamp, K. E. (2005). Sexually dimorphic patterns of space use throughout ontogeny in the spotted hyena (*Crocuta crocuta*). *Journal of Zoology*, 267(03), 271–281.

Boydston, E. E., Kapheim, K. M., Watts, H. E., Szykman, M., & Holekamp, K. E. (2003). Altered behaviour in spotted hyenas associated with increased human activity. *Animal Conservation*, 6(3), 207–219. https://doi.org/10.1017/S1367943003003263

Bray, E. E., MacLean, E. L., & Hare, B. A. (2015). Increasing arousal enhances inhibitory control in calm but not excitable dogs. *Animal Cognition*, 18(6), 1317–1329. https://doi.org/10.1007/s10071-015-0901-1

Bray, E. E., Maclean, E. L., & Hare, B. a. (2014). Context specificity of inhibitory control in dogs. *Animal Cognition*, 17(1), 15–31. https://doi.org/10.1007/s10071-013-0633-z

Bray, E. E., Sammel, M. D., Seyfarth, R. M., Serpell, J. A., & Cheney, D. L. (2017). Temperament and problem solving in a population of adolescent guide dogs. *Animal Cognition*. https://doi.org/10.1007/s10071-017-1112-8

Broadway, M. S., Samuelson, M. M., Christopher, J. L., Jett, S. E., & Lyn, H. (2017). Does size really matter? Investigating cognitive differences in spatial memory ability based on size in domestic dogs. *Behavioural Processes*, 138, 7–14. https://doi.org/http://doi.org/10.1016/j.beproc.2017.01.012

Brooks, Mollie, E., Kristensen, K., Benthem, Koen, J., V., Magnusson, A., Berg, Casper, W., Nielsen, A., … Bolker, Benjamin, M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2), 378. https://doi.org/10.32614/RJ-2017-066

Brucks, D., Marshall-Pescini, S., & Range, F. (2019). Dogs and wolves do not differ in their inhibitory control abilities in a non-social test battery. *Animal Cognition*, 22(1), 1–15. https://doi.org/10.1007/s10071-018-1216-9

Brucks, D., Marshall-Pescini, S., Wallis, L. J., Huber, L., & Range, F. (2017). Measures of Dogs' Inhibitory Control Abilities Do Not Correlate across Tasks. *Frontiers in Psychology*, 8, 849. https://doi.org/10.3389/fpsyg.2017.00849

Brust, V., Wuerz, Y., & Krüger, O. (2013). Behavioural flexibility and personality in zebra finches. *Ethology*, 119(7), 559–569. https://doi.org/10.1111/eth.12095

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, 40, e195. https://doi.org/10.1017/S0140525X16000959

Burnham, K., & Anderson, D. (2002). *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach*. (2nd ed.). New York: Springer.

Byrne, R. W. (1995a). Machiavellian Intelligence. In *The Thinking Ape: Evolutionary Origins of Intelligence* (pp. 195–209). New York: Oxford University Press Inc.

Byrne, R. W. (1995b). The thinking ape: Evolutionary origins of intelligence. New York: Oxford University Press.

Byrne, R. W. (1997). The Technical Intelligence hypothesis: An additional evolutionary stimulus to intelligence? In A Whiten & R. W. Byrne (Eds.), *Machiavellian Intelligence II: Extensions and Evaluations* (pp. 289-311). Cambridge, United Kingdom: Cambridge University Press.

Byrne, R. W., & Bates, L. A. (2007). Sociality, evolution and cognition. *Current Biology*, 17(16), R714-23. https://doi.org/10.1016/j.cub.2007.05.069

Byrne, R. W., & Whiten, A. (Eds.). (1988). *Machiavellian Intelligence: Social Expertise And The Evolution Of Intellect In Monkeys, Apes, And Humans*. Oxford University Press, USA.

Carere, C., & Locurto, C. (2011). Interaction between animal personality and animal cognition. *Current Zoology*, 57(4), 491–498. https://doi.org/10.1093/czoolo/57.4.491

Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development? *Trends in Cognitive Sciences*, 9(3), 104–110. https://doi.org/10.1016/j.tics.2005.01.011

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245–276.

Cauchard, L., Angers, B., Boogert, N. J., Lenarth, M., Bize, P., & Doligez, B. (2017). An experimental test of a causal link between problem-solving performance and reproductive success in wild great tits. *Frontiers in Ecology and Evolution*, 5(SEP), 1–8. https://doi.org/10.3389/fevo.2017.00107

Cauchard, L., Boogert, N. J., Lefebvre, L., Dubois, F., & Doligez, B. (2013). Problem-solving performance is correlated with reproductive success in a wild bird population. *Animal Behaviour*, 85(1), 19–26. https://doi.org/10.1016/j.anbehav.2012.10.005

Cauchoix, M., Chow, P. K. Y., van Horik, J. O., Atance, C. M., Barbeau, E. J., Barragan-Jason, G., … Morand-Ferron, J. (2018). The repeatability of cognitive performance: a meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756). http://dx.doi.org/10.1098/rstb.2017.0281

Cauchoix, M., Hermer, E., Chaine, A. S., & Morand-Ferron, J. (2017). Cognition in the field: comparison of reversal learning performance in captive and wild passerines. *Scientific Reports*, 7(1), 12945. https://doi.org/10.1038/s41598-017-13179-5

Chen, J., Zou, Y., Sun, Y.-H., & ten Cate, C. (2019). Problem-solving males become more attractive to female budgerigars. *Science*, 363(6423), 166–167. https://doi.org/10.1126/science.aau8181

Chittka, L., Skorupski, P., & Raine, N. E. (2009). Speed–accuracy tradeoffs in animal decision making. *Trends in Ecology & Evolution*, 24(7), 400–407. https://doi.org/10.1016/j.tree.2009.02.010

Chow, P. K. Y., Lea, S. E. G., & Leaver, L. A. (2016). How practice makes perfect: the role of persistence, flexibility and learning in problem-solving efficiency. *Animal Behaviour*, 112(Supplement C), 273–283. https://doi.org/10.1016/j.anbehav.2015.11.014

Clark, F. E. (2017). Cognitive enrichment and welfare: Current approaches and future directions. *Animal Behavior and Cognition*, 4(1), 52–71. https://doi.org/10.12966/abc.05.02.2017

Clutton-Brock, T. H., & Harvey, P. H. (1980). Primates, brains and ecology. *Journal of Zoology*, 190(3), 309–323. https://doi.org/10.1111/j.1469-7998.1980.tb01430.x

Cole, E. F., Cram, D. L., & Quinn, J. L. (2011). Individual variation in spontaneous problem-solving performance among wild great tits. *Animal Behaviour*, 81(2), 491–498. https://doi.org/10.1016/j.anbehav.2010.11.025

Cole, E. F., Morand-Ferron, J., Hinks, A. E., & Quinn, J. L. (2012). Cognitive ability influences reproductive life history variation in the wild. *Current Biology*, 22(19), 1808–1812. https://doi.org/10.1016/j.cub.2012.07.051

Cole, E. F., Quinn, J. L., Cole, E. F., & Quinn, J. L. (2012). Personality and problem-solving performance explain competitive ability in the wild. *Proceedings of the Royal Society B: Biological Sciences,* 279(1731), 1168–1175. https://doi.org/10.1098/rspb.2011.1539

Cooper, S. M., Holekamp, K. E., & Smale, L. (1999). A seasonal feast: long-term analysis of feeding behaviour in the spotted hyaena (*Crocuta crocuta*). *African Journal of Ecology*, 37(2), 149–160. https://doi.org/10.1046/j.1365-2028.1999.00161.x

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences*, 10(3), 131–138. https://doi.org/10.1016/j.tics.2006.01.007

Damerius, L. A., Burkart, J. M., van Noordwijk, M. A., Haun, D. B. M. M., Kosonen, Z. K., Galdikas, B. M. F. F., … van Schaik, C. P. (2018). General cognitive abilities in orangutans (*Pongo abelii and Pongo pygmaeus*). *Intelligence*, 74(October 2018), 3–11. https://doi.org/10.1016/j.intell.2018.10.008

Daniels, S. E., Fanelli, R. E., Gilbert, A., & Benson-Amram, S. (2019). Behavioral flexibility of a generalist carnivore. *Animal Cognition*, 22(3), 387–396. https://doi.org/10.1007/s10071-019-01252-7

de Waal, F. B. M. (2013). *The bonobo and the atheist: In search of humanism among the primates*. New York, NY: W. W. Norton & Company.

Deaner, R. O., Isler, K., Burkart, J., & van Schaik, C. (2007). Overall brain size, and not encephalization quotient, best predicts cognitive ability across non-human primates. *Brain, Behavior and Evolution*, 70(2), 115–124.

Deaner, R. O., Nunn, C. L., & van Schaik, C. P. (2000). Comparative tests of primate cognition: different scaling methods produce different results. *Brain, Behavior and Evolution*, 55(1), 44–52.

DeCasien, A. R., Williams, S. A., & Higham, J. P. (2017). Primate brain size is predicted by diet but not sociality. *Nature Ecology & Evolution*, 1, 0112. https://doi.org/10.1038/s41559-017-0112

Diamond, A. (1990). Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of, Inhibitory Control in Reaching. *Annals of the New York Academy of Sciences*, 608(1), 637–676.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diquelou, M. C., Griffin, A. S., & Sol, D. (2015). The role of motor diversity in foraging innovations: a cross-species comparison in urban birds. *Behavioral Ecology*, 27(2), 584–591.

Drea, C. M., Hawk, J. E., & Glickman, S. E. (1997). The Emergence of Affiliative Behavior in Infant Spotted Hyenas (*Crocuta crocuta*). *Annals of the New York Academy of Sciences*, 807(1), 498–500.

Ducatez, S., Audet, J.-N., & Lefebvre, L. (2019). Speed–accuracy trade-off, detour reaching and response to PHA in Carib grackles. *Animal Cognition*, 0(0), 0. https://doi.org/10.1007/s10071-019-01258-1

Ducatez, S., Audet, J. N., & Lefebvre, L. (2015). Problem-solving and learning in Carib grackles: individuals show a consistent speed–accuracy trade-off. *Animal Cognition*, 18(2), 485–496.

Ducatez, S., Clavel, J., & Lefebvre, L. (2014). Ecological generalism and behavioural innovation in birds: technical intelligence or the simple incorporation of new foods? *The Journal of Animal Ecology*. https://doi.org/10.1111/1365-2656.12255

Dunbar, R. I. M. (1998). The social brain hypothesis. Evolutionary Anthropology: Issues, News, and Reviews, 6(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.3.CO;2-P

Dunbar, R. I. M., & Shultz, S. (2007). Evolution in the social brain. *Science*. 317(5843), 1344–1347. https://doi.org/10.1126/science.1145463

Dunbar, R. I. M., & Shultz, S. (2017). Why are there so many explanations for primate brain evolution? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1727), 20160244. http://dx.doi.org/10.1098/rstb.2016.0244

Emery, N. J., Seed, A. M., von Bayern, A. M. P., & Clayton, N. S. (2007). Cognitive adaptations of social bonding in birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 489–505. https://doi.org/10.1098/rstb.2006.1991

Engh, A. L., Esch, K., Smale, L., & Holekamp, K. (2000). Mechanisms of maternal rank "inheritance" in the spotted hyaena, *Crocuta crocuta*. *Animal Behaviour*, 60(3), 323–332. https://doi.org/10.1006/anbe.2000.1502

Fagnani, J., Barrera, G., Carballo, F., & Bentosela, M. (2016). Is previous experience important for inhibitory control? A comparison between shelter and pet dogs in A-not-B and cylinder tasks. *Animal Cognition*, 1–8. https://doi.org/https://doi.org/10.1007/s10071-016-1024-z

Federspiel, I. G., Garland, A., Guez, D., Bugynar, T., Healy, S. D., Güntürkün, O., & Griffin, A. S. (2017). Adjusting foraging strategies: a comparison of rural and urban common mynas (*Acridotheres tristis*). *Animal Cognition*, 20(1), 65–74. https://doi.org/10.1007/s10071-016-1045-7

Fedorova, N., Evans, C. L., & Byrne, R. W. (2017). Living in stable social groups is associated with reduced brain size in woodpeckers (*Picidae*). *Biology Letters*, 13(3), 20170008. http://dx.doi.org/10.1098/rsbl.2017.0008

Feyereisen, P., & Charlot, V. (2008). Are There Uniform Age-Related Changes Across Tasks Involving Inhibitory Control Through Access, Deletion, and Restraint Functions? A Preliminary Investigation. *Experimental Aging Research*, 34(4), 392–418. https://doi.org/10.1080/03610730802271880

Frank, L. G. (1986). Social organization of the spotted hyaena *Crocuta crocuta*. II. Dominance and reproduction. *Animal Behaviour*, 34(5), 1510–1527. http://dx.doi.org/10.1016/S0003-3472(86)80221-4

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Fristoe, T. S., & Botero, C. A. (2019). Alternative ecological strategies lead to avian brain size bimodality in variable habitats. *Nature Communications*, 10(1), 3818. https://doi.org/10.1038/s41467-019-11757-x

Fristoe, T. S., Iwaniuk, A. N., & Botero, C. A. (2017). Big brains stabilize populations and facilitate colonization of variable habitats in birds. *Nature Ecology & Evolution*, 1(11), 1706. https://doi.org/10.1038/s41559-017-0316-2

Gade, D. W. (2006). Hyenas and humans in the Horn of Africa. *Geographical Review*, 96(4), 609–632.

Gosling, S. D., Hawk, J. E., Beer, J. S., Block, J., John, O. P., King, J. E., & Mcfarling, U. L. (1998). Personality Dimensions in Spotted Hyenas (*Crocuta crocuta*). *Journal of Comparative Psychology*, 112(2), 107–118. https://doi.org/10.1037/0735-7036.112.2.107

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79–132.

Green, D. S., & Holekamp, K. E. (2019). Pastoralist activities affect the movement patterns of a large African carnivore, the spotted hyena (*Crocuta crocuta*). *Journal of Mammalogy*, 100(6), 1941–1953. https://doi.org/10.1093/jmammal/gyz135

Green, D. S., Johnson-Ulrich, L., Couraud, H. E., & Holekamp, K. E. (2018). Anthropogenic disturbance induces opposing population trends in spotted hyenas and African lions. *Biodiversity and Conservation*, 27(4), 871–889. https://doi.org/10.1007/s10531-017-1469-7

Greenberg, J. R., & Holekamp, K. E. (2017). Human disturbance affects personality development in a wild carnivore. *Animal Behaviour*, 132, 303–312. https://doi.org/10.1016/j.anbehav.2017.08.023

Greenough, W. T., Black, J. E., & Wallace, C. S. (1993). Experience and brain development. In M. H. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain development and cognition: a reader* (Vol. 2, pp. 186–216). Oxford, England: Blackwell Publishers Ltd.

Greggor, A. L., Thornton, A., & Clayton, N. S. (2015). Neophobia is not only avoidance: improving neophobia tests by combining cognition and ecology. *Current Opinion in Behavioral Sciences*, 6, 82–89. https://doi.org/10.1016/j.cobeha.2015.10.007

Griffin, A. S. (2016). Innovativeness as an emergent property: a new alignment of comparative and experimental research on animal innovation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1690), 20150544. https://doi.org/10.1098/rstb.2015.0544

Griffin, A. S., & Diquelou, M. C. (2015). Innovative problem solving in birds: a cross-species comparison of two highly successful passerines. *Animal Behaviour*, 100, 84–94. https://doi.org/10.1016/j.anbehav.2014.11.012

Griffin, A. S., & Guez, D. (2014). Innovation and problem solving: A review of common mechanisms. *Behavioural Processes*, 109(Pt B), 121–134. https://doi.org/10.1016/j.beproc.2014.08.027

Griffin, A. S., & Guez, D. (2015). Innovative problem solving in nonhuman animals: the effects of group size revisited. *Behavioral Ecology*, 26(3), 722–734.

Griffin, A. S., Guez, D., Lermite, F., & Patience, M. (2013). Tracking changing environments: innovators are fast, but not flexible learners. PloS One, 8(12), e84907.

Griffin, A. S., Guillette, L. M., & Healy, S. D. (2015). Cognition and personality: an analysis of an emerging field. *Trends in Ecology & Evolution*, 30(4), 207–214. https://doi.org/10.1016/j.tree.2015.01.012

Griffin, A. S., Netto, K., & Peneaux, C. (2017). Neophilia, innovation and learning in an urbanized world: a critical evaluation of mixed findings. *Current Opinion in Behavioral Sciences*, 16, 15–22. http://dx.doi.org/10.1016/j.cobeha.2017.01.004

Guillette, L. M., Naguib, M., & Griffin, A. S. (2017). Individual differences in cognition and personality. *Behavioural Processes*, 134, 1–3. https://doi.org/10.1016/j.beproc.2016.12.001

Hartig, F. (2019). DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. Retrieved from http://florianhartig.github.io/DHARMa/

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, 26(4), 433–443.

Harvey, P. H., & Krebs, J. R. (1990). Comparing Brains. *Science*, 249(4965), 140–146.

Hayward, M. W. (2006). Prey preferences of the spotted hyaena (*Crocuta crocuta*) and degree of dietary overlap with the lion (*Panthera leo*). *Journal of Zoology*, 270(4), 606–614. https://doi.org/10.1111/j.1469-7998.2006.00183.x

Healy, S. D, & Rowe, C. (2007). A critique of comparative studies of brain size. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 453–464. https://doi.org/10.1098/rspb.2006.3748

Healy, Susan D., & Rowe, C. (2013). Costs and benefits of evolving a larger brain: doubts over the evidence that large brains lead to better cognition. *Animal Behaviour*, 86(4), e1–e3. https://doi.org/10.1016/j.anbehav.2013.05.017

Healy, Susan D, & Rowe, C. (2014). Animal cognition in the wild. *Behavioural Processes*, 109 Pt B, 101–102. https://doi.org/10.1016/j.beproc.2014.11.013

Herculano-Houzel, S. (2017). Numbers of neurons as biological correlates of cognitive capability. *Current Opinion in Behavioral Sciences*, 16, 1–7. http://dx.doi.org/10.1016/j.cobeha.2017.02.004

Herculano-Houzel, S. (2011). Brains matter, bodies maybe not: the case for examining neuron numbers irrespective of body size. *Annals of the New York Academy of Sciences*, 1225(1), 191–199. https://doi.org/10.1111/j.1749-6632.2011.05976.x

Herczeg, G., Urszán, T. J., Orf, S., Nagy, G., Kotrschal, A., & Kolm, N. (2019). Brain size predicts behavioural plasticity in guppies (*Poecilia reticulata*): An experiment. *Journal of Evolutionary Biology*, 32(3), 218–226. https://doi.org/10.1111/jeb.13405

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science*, 317(5843), 1360–1366. https://doi.org/10.1126/science.1146282

Hofer, H., & East, M. L. (1993). The commuting system of Serengeti spotted hyaenas: how a predator copes with migratory prey. I. Social organization. *Animal Behaviour*, 46(3), 547–557. https://doi.org/10.1006/anbe.1993.1222

Holekamp, K. E. (2007). Questioning the social intelligence hypothesis. *Trends in Cognitive Sciences*, 11(2), 65–69. https://doi.org/10.1016/j.tics.2006.11.003

Holekamp, K. E., & Benson-Amram, S. (2017). The evolution of intelligence in mammalian carnivores. *Interface Focus*, 7(3), 20160108. https://doi.org/10.1098/rsfs.2016.0108

Holekamp, K. E., Dantzer, B., Stricker, G., Shaw Yoshida, K. C., & Benson-Amram, S. (2015). Brains, brawn and sociality: a hyaena's tale. *Animal Behaviour*, 103, 237–248. https://doi.org/10.1016/j.anbehav.2015.01.023

Holekamp, K. E., & Dloniak, S. M. (2010). Intraspecific Variation in the Behavioral Ecology of a Tropical Carnivore, the Spotted Hyena. In *Behavioral ecology of tropical animals* (1st ed., Vol. 42, pp. 189–229). Elsevier. https://doi.org/10.1016/S0065-3454(10)42006-9

Holekamp, K. E., Sakai, S. T., & Lundrigan, B. L. (2007). The spotted hyena (*Crocuta crocuta*) as a model system for study of the evolution of intelligence. *Journal of Mammalogy*, 88(3), 545–554. https://doi.org/10.1644/06-MAMM-S-361R1.1

Holekamp, K. E., & Smale, L. (1991). Dominance acquisition during mammalian social development: the "inheritance" of maternal rank. *American Zoologist*, 31(2), 306–317. https://doi.org/10.1093/icb/31.2.306

Holekamp, K. E., Smale, L., Berg, R., & Cooper, S. M. (1997). Hunting rates and hunting success in the spotted hyena (*Crocuta crocuta*). *Journal of Zoology*, 242(1), 1–15. https://doi.org/10.1111/j.1469-7998.1997.tb02925.x

Holekamp, K. E., Smale, L., & Szykman, M. (1996). Rank and reproduction in the female spotted hyaena. *Reproduction*, 108(2), 229–237. https://doi.org/10.1530/jrf.0.1080229

Holekamp, K. E., Smith, J. E., Strelioff, C. C., Van Horn, R. C., & Watts, H. E. (2012). Society, demography and genetic structure in the spotted hyena. *Molecular Ecology*, 21(3), 613–632. https://doi.org/10.1111/j.1365-294X.2011.05240.x

Horschler, D. J., Hare, B., Call, J., Kaminski, J., Miklósi, Á., & MacLean, E. L. (2019). Absolute brain size predicts dog breed differences in executive function. *Animal Cognition*, 1–12. https://doi.org/https://doi.org/10.1007/s10071-018-01234-1

Huebner, F., & Fichtel, C. (2015). Innovation and behavioral flexibility in wild redfronted lemurs (*Eulemur rufifrons*). *Animal Cognition*, 18(3), 777–787. https://doi.org/10.1007/s10071-015-0844-6

Huebner, F., Fichtel, C., & Kappeler, P. M. (2018). Linking cognition with fitness in a wild primate: fitness correlates of problem-solving performance and spatial learning ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756), 20170295. https://doi.org/10.1098/rstb.2017.0295

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303–317). Cambridge, United Kingdom: Cambridge University Press.

Isden, J., Panayi, C., Dingle, C., & Madden, J. (2013). Performance in cognitive and problem-solving tasks in male spotted bowerbirds does not correlate with mating success. *Animal Behaviour*, 86(4), 829–838. https://doi.org/10.1016/j.anbehav.2013.07.024

Isler, K., & van Schaik, C. P. (2009). The expensive brain: a framework for explaining evolutionary changes in brain size. *Journal of Human Evolution*, 57(4), 392–400. https://doi.org/10.1016/j.jhevol.2009.04.009

Jelbert, S. A., Taylor, A. H., & Gray, R. D. (2016). Does absolute brain size really predict self-control? Hand-tracking training improves performance on the A-not-B task. *Biology Letters*, 12(2), 20150871-. https://doi.org/10.1098/rsbl.2015.0871

Johnson-Ulrich, L., Benson-Amram, S., & Holekamp, K. E. (2019). Fitness Consequences of Innovation in Spotted Hyenas. *Frontiers in Ecology and Evolution*, 7, 443. https://doi.org/10.3389/fevo.2019.00443

Johnson-Ulrich, L., & Holekamp, K. E. (2020). Group size and social rank predict inhibitory control in spotted hyaenas. *Animal Behaviour*, 160, 157–168. https://doi.org/10.1016/j.anbehav.2019.11.020

Johnson-Ulrich, L., Johnson-Ulrich, Z., & Holekamp, K. (2018). Proactive behavior, but not inhibitory control, predicts repeated innovation by spotted hyenas tested with a multi-access box. *Animal Cognition*, 21(3), 379–392. https://doi.org/10.1007/s10071-018-1174-2

Johnson, M. H., Munakata, Y., & Gilmore, R. O. (Eds.). (2005). *Brain Development and Cognition* (2nd edition). Oxford, UK: Blackwell Publishers Ltd. https://doi.org/10.1002/9780470753507

Jolly, A. (1966). Lemur Social Behavior and Primate Intelligence. *Science*, 153(3735), 501–506. https://doi.org/10.1126/science.153.3735.501

Jones, C. B. (2005). *Behavioral Flexibility in Primates: Causes and Consequences*. New York: Springer.

Kabadayi, C., Bobrowicz, K., & Osvath, M. (2018). The detour paradigm in animal cognition. *Animal Cognition*, 21(1), 21–35. https://doi.org/10.1007/s10071-017-1152-0

Kabadayi, C., Krasheninnikova, A., O'Neill, L., van de Weijer, J., Osvath, M., & von Bayern, A. M. P. (2017). Are parrots poor at motor self-regulation or is the cylinder task poor at measuring it? *Animal Cognition*, 20(6), 1137–1146. https://doi.org/10.1007/s10071-017-1131-5

Kabadayi, C., Taylor, L. A., von Bayern, A. M. P., & Osvath, M. (2016). Ravens, New Caledonian crows and jackdaws parallel great apes in motor self-regulation despite smaller brains. *Royal Society Open Science*, 3(4), 160104. https://doi.org/10.1098/rsos.160104

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151.

Kamil, A. (1987). A synthetic approach to the study of animal intelligence. *Nebraska Symposium on Motivation*, 35, 257–308.

Kang, F., Goulet, C. T., & Chapple, D. G. (2017). The impact of urbanization on learning ability in an invasive lizard. *Biological Journal of the Linnean Society*, 123(1), 55–62.

Kappeler, P. M. (2019). A framework for studying social complexity. *Behavioral Ecology and Sociobiology*, 73(1), 13. https://doi.org/10.1007/s00265-018-2601-8

Kark, S., Iwaniuk, A., Schalimtzek, A., & Banker, E. (2007). Living in the city: can anyone become an "urban exploiter"? *Journal of Biogeography*, 34(4), 638–651. https://doi.org/10.1111/j.1365-2699.2006.01638.x

Kassambara, A., & Kosinski, M. (2018). Drawing survival curves using "ggplot2." Retrieved from https://cran.r-project.org/package=survminer

Keagy, J., Savard, J.-F., & Borgia, G. (2009). Male satin bowerbird problem-solving ability predicts mating success. *Animal Behaviour*, 78(4), 809–817. https://doi.org/10.1016/j.anbehav.2009.07.011

King, B. J. (1986). Extractive foraging and the evolution of primate intelligence. *Human Evolution*, 1(4), 361–372. https://doi.org/10.1007/BF02436709

Kolb, B., & Gibb, R. (2011). Brain plasticity and behaviour in the developing brain. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 20(4), 265.

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., … Kolm, N. (2013a). Artificial Selection on Relative Brain Size in the Guppy Reveals Costs and Benefits of Evolving a Larger Brain. *Current Biology*, 23(2), 168–171. https://doi.org/10.1016/j.cub.2012.11.058

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., … Kolm, N. (2013b). The benefit of evolving a larger brain: big-brained guppies perform better in a cognitive task. *Animal Behaviour*, 86(4), e4–e6. https://doi.org/10.1016/j.anbehav.2013.07.011

Kruuk, H. (1972). *The spotted hyena: a study of predation and social behavior*. Brattleboro, VT: Echo Point Books & Media.

Kudo, H., & Dunbar, R. I. M. (2001). Neocortex size and social network size in primates. *Animal Behaviour*, 62(4), 711–722. https://doi.org/10.1006/anbe.2001.1808

Kummer, H., & Goodall, J. (1985). Conditions of Innovative Behaviour in Primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 308(1135), 203–214. https://doi.org/10.1098/rstb.1985.0020

Langbein, J., Siebert, K., & Nürnberg, G. (2009). On the use of an automated learning device by group-housed dwarf goats: Do goats seek cognitive challenges? *Applied Animal Behaviour Science*, 120(3–4), 150–158. https://doi.org/10.1016/j.applanim.2009.07.006

Lea, S. E. G., Chow, P. K. Y., Leaver, L. A., & McLaren, I. P. L. (2020). Behavioral flexibility: A review, a model, and some exploratory tests. *Learning & Behavior*. https://doi.org/10.3758/s13420-020-00421-w

Lefebvre, L., Reader, S. M., & Sol, D. (2004). Brains, innovations and evolution in birds and primates. *Brain, Behavior and Evolution*, 63(4), 233–246. https://doi.org/10.1159/000076784

Lefebvre, L., Reader, S. M., & Sol, D. (2013). Innovating innovation rate and its relationship with brains, ecology and general intelligence. *Brain, Behavior and Evolution*, 81(3), 143–145. https://doi.org/10.1159/000348485

Lefebvre, L., Whittle, P., Lascaris, E., & Finkelstein, A. (1997). Feeding innovations and forebrain size in birds. *Animal Behaviour*, 53(3), 549–560. https://doi.org/10.1006/anbe.1996.0330

Lermite, F., Peneaux, C., & Griffin, A. S. (2017). Personality and problem-solving in common mynas (*Acridotheres tristis*). *Behavioural Processes*, 134, 87–94. https://doi.org/10.1016/j.beproc.2016.09.013

Lewin, N., Swanson, E. M., Williams, B. L., & Holekamp, K. E. (2017). Juvenile concentrations of IGF-1 predict life-history trade-offs in a wild mammal. *Functional Ecology*, 31(4), 894–902. https://doi.org/10.1111/1365-2435.12808

Liao, W. B., Lou, S. L., Zeng, Y., Kotrschal, A., Leips, J., & Bronstein, J. L. (2016). Large Brains, Small Guts: The Expensive Tissue Hypothesis Supported within Anurans. *The American Naturalist*, 188(6), 0.

Lindsay, W. R., Houck, J. T., Giuliano, C. E., & Day, L. B. (2015). Acrobatic courtship display coevolves with brain size in manakins (*Pipridae*). *Brain, Behavior and Evolution*, 85(1), 29–36. https://doi.org/10.1159/000369244

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., … Zhao, Y. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), E2140-8. https://doi.org/10.1073/pnas.1323533111

MacLean, E. L., Sandel, A. A., Bray, J., Oldenkamp, R. E., Reddy, R. B., & Hare, B. A. (2013). Group size predicts social but not nonsocial cognition in lemurs. *PLoS One*, 8(6), e66359. https://dx.plos.org/10.1371/journal.pone.0066359

Maklakov, A. A., Immler, S., Gonzalez-Voyer, A., Rönn, J., & Kolm, N. (2011). Brains and the city: big-brained passerine birds succeed in urban environments. *Biology Letters*, 7(5), 730–732. https://doi.org/10.1098/rsbl.2011.0341

Manrique, H. M., Völter, C. J., & Call, J. (2013). Repeated innovation in great apes. *Animal Behaviour*, 85(1), 195–202. https://doi.org/10.1016/j.anbehav.2012.10.026

Marshall-Pescini, S., Virányi, Z., & Range, F. (2015). The effect of domestication on inhibitory control: Wolves and dogs Compared. *PLoS One*, 10(2), e0118469. https://doi.org/10.1371/journal.pone.0118469

Martin, P., & Bateson, P. (1993). *Measuring behaviour*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9781139168342

Matzel, L. D., Grossman, H., Light, K., Townsend, D., & Kolata, S. (2008). Age-related declines in general cognitive abilities of Balb/C mice are associated with disparities in working memory, body weight, and general activity. *Learning & Memory*, 15(10), 733–746.

McCune, K. B., Jablonski, P., Lee, S., & Ha, R. R. (2019). Captive jays exhibit reduced problem-solving performance compared to wild conspecifics. *Royal Society Open Science*, 6(1), 181311. https://doi.org/10.1098/rsos.181311

McGowan, R. T. S., Robbins, C. T., Alldredge, J. R., & Newberry, R. C. Contrafreeloading in grizzly bears: implications for captive foraging enrichment. *Zoo Biology*, 29(4), 484–502. https://doi.org/10.1002/zoo.20282

Menzel, E. (1974). A group of chimpanzees in a one-acre field. In A. Shrier & F. Stollnitz (Eds.), *Behavior of non-human primates* (pp. 83–153). San Diego: Academic Press.

Milton, K. (1981). Distribution Patterns of Tropical Plant Foods as an Evolutionary Stimulus to Primate Mental Development. *American Anthropologist*, 83(3), 534–548. https://doi.org/10.1525/aa.1981.83.3.02a00020

Mischel, W., Shoda, Y., & Rodriguez, M. (1989). Delay of gratification in children. *Science*, 244(4907), 933–938. https://doi.org/10.1126/science.2658056

Mitchell, C. (2016). The Evolution of Brains and Cognitive Abilities. In P. Pontarotti (Ed.), *Evolutionary Biology* (pp. 73–87). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-41324-2_5

Møller, A. P. (2009). Successful city dwellers: a comparative study of the ecological characteristics of urban birds in the Western Palearctic. *Oecologia*, 159(4), 849–858. https://doi.org/10.1007/s00442-008-1259-8

Morand-Ferron, J., Cole, E. F., & Quinn, J. L. (2015). Studying the evolutionary ecology of cognition in the wild: a review of practical and conceptual challenges. *Biological Reviews of the Cambridge Philosophical Society*. https://doi.org/10.1111/brv.12174

Müller, C. A., Riemer, S., Virányi, Z., Huber, L., & Range, F. (2016). Inhibitory Control, but Not Prolonged Object-Related Experience Appears to Affect Physical Problem-Solving Performance of Pet Dogs. *PLoS One*, 11(2), e0147753. https://doi.org/10.1371/journal.pone.0147753

Muthukrishna, M., Doebeli, M., Chudek, M., & Henrich, J. (2018). The Cultural Brain Hypothesis: How culture drives brain expansion, sociality, and life history. *PloS Computational Biology*, 14(11), e1006504. https://doi.org/10.1371/journal.pcbi.1006504

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews*, 85(4), 935–956. https://doi.org/10.1111/j.1469-185X.2010.00141.x

Nave, G., Jung, W. H., Karlsson Linnér, R., Kable, J. W., & Koellinger, P. D. (2019). Are Bigger Brains Smarter? Evidence From a Large-Scale Preregistered Study. *Psychological Science*, 30(1), 43–54. https://doi.org/10.1177/0956797618808470

Overington, S. E., Griffin, A. S., Sol, D., & Lefebvre, L. (2011). Are innovative species ecological generalists? A test in North American birds. *Behavioral Ecology*, 22(6), 1286–1293. https://doi.org/10.1093/beheco/arr130

Overington, Sarah E, Cauchard, L., Côté, K.-A., & Lefebvre, L. (2011). Innovative foraging behaviour in birds: What characterizes an innovator? *Behavioural Processes*, 87(3), 274–285. https://doi.org/10.1016/j.beproc.2011.06.002

Papp, S., Vincze, E., Preiszner, B., Liker, A., & Bókony, V. (2014). A comparison of problem-solving success between urban and rural house sparrows. *Behavioral Ecology and Sociobiology*, 69(3), 471–480. https://doi.org/10.1007/s00265-014-1859-8

Parker, S. T. (2015). Re-evaluating the extractive foraging hypothesis. *New Ideas in Psychology*, 37, 1–12. https://doi.org/10.1016/j.newideapsych.2014.11.001

Partecke, J., Schwabl, I., & Gwinner, E. (2006). Stress and the city: urbanization and its effects on the stress physiology in European blackbirds. *Ecology*, 87(8), 1945–1952. https://doi.org/10.1890/0012-9658(2006)87[1945:SATCUA]2.0.CO;2

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373–1379. https://doi.org/10.1016/s0895-4356(96)00236-3

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry*, 20(1), 98–108. https://doi.org/10.1038/mp.2014.105

Powell, L. E., Isler, K., & Barton, R. A. (2017). Re-evaluating the link between brain size and behavioural ecology in primates. *Proceedings of the Royal Society B: Biological Sciences*, 284(1865). http://dx.doi.org/10.1098/rspb.2017.1765

Prasher, S., Evans, J. C., Thompson, M. J., & Morand-Ferron, J. (2019). Characterizing innovators: Ecological and individual predictors of problem-solving performance. *PloS One*, 14(6), e0217464. https://doi.org/10.1371/journal.pone.0217464

Preiszner, B., Papp, S., Pipoly, I., Seress, G., Vincze, E., Liker, A., & Bókony, V. (2017). Problem-solving performance and reproductive success of great tits in urban and forest habitats. *Animal Cognition*, 20(1), 53–63. https://doi.org/10.1007/s10071-016-1008-z

Pritchard, D. J., Hurly, T. A., Tello-Ramos, M. C., & Healy, S. D. (2016). Why study cognition in the wild (and how to test it)? *Journal of the Experimental Analysis of Behavior*, 105(1), 41–55. https://doi.org/10.1002/jeab.195

Quinn, J. L., Cole, E. F., Reed, T. E., & Morand-Ferron, J. (2016). Environmental and genetic determinants of innovativeness in a natural population of birds. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1690), 20150184. https://doi.org/10.1098/rstb.2015.0184

R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Ramsey, G., Bastian, M. L., & van Schaik, C. (2007). Animal innovation defined and operationalized. *Behavioral and Brain Sciences*, 30(4), 393–407. https://doi.org/10.1017/S0140525X07002373

Reader, S. M., Hager, Y., Laland, K. N. K., Munch, S. B., Tebbich, S., Bshary, R., … Wheeler, P. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366(1567), 1017–1027. https://doi.org/10.1098/rstb.2010.0342

Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences of the United States of America*, 99(7), 4436–4441. https://doi.org/10.1073/pnas.062041299

Reader, S. M., & Laland, K. N. (Eds.). (2003). *Animal Innovation*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198526223.001.0001

Reader, S. M., Morand-Ferron, J., & Flynn, E. (2016). Animal and human innovation: novel problems and novel solutions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1690), 20150182. https://doi.org/10.1098/rstb.2015.0182

Reddy, R. B., MacLean, E. L., Sandel, A. A., & Hare, B. (2015). Social inhibitory control in five lemur species. *Primates*, 56(3), 241–252. https://doi.org/10.1007/s10329-015-0467-1

Ricklefs, R. E., & Wikelski, M. (2002). The physiology/life-history nexus. *Trends in Ecology and Evolution*, 17(10), 462–468. https://doi.org/10.1016/S0169-5347(02)02578-8

Rosati, A. G. (2017). Foraging Cognition: Reviving the Ecological Intelligence Hypothesis. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2017.05.011

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, 9(5), 250–257. https://doi.org/10.1016/j.tics.2005.03.005

Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. *Behavioral Ecology*, 25(6), 1287–1292. https://doi.org/10.1093/beheco/aru090

Rudolph, K., & Fichtel, C. (2017). Inhibitory control in douc langurs (*Pygathrix nemaeus and P. cinerea*). *Vietnamese Journal of Primatology*, 2(5), 73–81.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, 30(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Santacà, M., Busatta, M., Savaşçı, B. B., Lucon-Xiccato, T., & Bisazza, A. (2019). The effect of experience and olfactory cue in an inhibitory control task in guppies, *Poecilia reticulata*. *Animal Behaviour,* 151, 1–7. https://doi.org/10.1016/j.anbehav.2019.03.003

Sayol, F., Maspons, J., Lapiedra, O., Iwaniuk, A. N., Székely, T., & Sol, D. (2016). Environmental variation and the evolution of large brains in birds. *Nature Communications*, 7, 13971. https://doi.org/10.1038/ncomms13971

Schramme, E. (2015). Social Structure of Spotted Hyena (*Crocuta crocuta*) Populations around Mekelle city in Tigray, Ethiopia (Master's thesis, Universiteit Antwerpen, Antwerp, Belgium). Available from Semantic Scholar (Corpus ID: 109929783).

Schuck-Paim, C., Alonso, W. J., & Ottoni, E. B. (2008). Cognition in an ever-changing world: climatic variability is associated with brain size in Neotropical parrots. *Brain, Behavior and Evolution*, 71(3), 200–215. https://doi.org/10.1159/000119710

Schuster, A. C., Zimmermann, U., Hauer, C., & Foerster, K. (2017). A behavioural syndrome, but less evidence for a relationship with cognitive traits in a spatial orientation context. *Frontiers in Zoology*, 14(1), 19. https://doi.org/10.1186/s12983-017-0204-2

Shaw, R. C. (2017). Testing cognition in the wild: factors affecting performance and individual consistency in two measures of avian cognition. *Behavioural Processes*, 134, 31–36. http://dx.doi.org/10.1016/j.beproc.2016.06.004

Shaw, R. C., & Schmelz, M. (2017). Cognitive test batteries in animal cognition research: evaluating the past, present and future of comparative psychometrics. *Animal Cognition*, 20(6), 1003–1018. https://doi.org/10.1007/s10071-017-1135-1

Shettleworth, S. J. (1998). *Cognition, evolution, and behavior*. Oxford: Oxford University Press.

Shuker, D. M., Barrett, L., Dickins, T. E., Scott-Phillips, T. C., & Barton, R. A. (2017). General intelligence does not help us understand cognitive evolution. *The Behavioral and Brain Sciences*, 40(1916), e218. https://doi.org/10.1017/S0140525X16001771

Shultz, S., Bradbury, R. B., L Evans, K., Gregory, R. D., & Blackburn, T. M. (2005). Brain size and resource specialization predict long-term population trends in British birds. *Proceedings of the Royal Society B: Biological Sciences*, 272(1578), 2305–2311. https://doi.org/10.1098/rspb.2005.3250

Shultz, S., & Dunbar, R. I. M. (2010). Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. *Journal of Comparative Psychology*, 124(3), 252–260. https://doi.org/10.1037/a0018894

Sih, A. (2013). Understanding variation in behavioural responses to human-induced rapid environmental change: a conceptual overview. *Animal Behaviour*, 85(5), 1077–1088. https://doi.org/10.1016/j.anbehav.2013.02.017

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004). Behavioral syndromes: an integrative overview. *The Quarterly Review of Biology*, 79(3), 241–277. https://doi.org/10.1086/422893

Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 2762–2772. https://doi.org/10.1098/rstb.2012.0216

Sikes, R. S. (2016). 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *Journal of Mammalogy*, 97(3), 663–688. https://doi.org/10.1093/jmammal/gyw078

Sinclair, A. R. E., & Norton-Griffiths, M. (Eds.). (1984). *Serengeti: dynamics of an ecosystem*. Chicago, IL: University of Chicago Press.

Sinervo, B., Svensson, E., & Comendant, T. (2000). Density cycles and an offspring quantity and quality game driven by natural selection. *Nature*, 406(6799), 985–988. https://doi.org/10.1038/35023149

Smaers, J. B., & Soligo, C. (2013). Brain reorganization, not relative brain size, primarily characterizes anthropoid brain evolution. *Proceedings of the Royal Society B: Biological Sciences*, 280(1759), 20130269. https://doi.org/10.1098/rspb.2013.0269

Smale, L., Nunes, S., & Holekamp, K. E. (1997). Sexually Dimorphic Dispersal in Mammals: Patterns, Causes, and Consequences. In P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon, & M. Milinski (Eds.), *Advances in the study of behaviour* (Vol. 26, pp. 181–250). New York, NY: Academic Press. https://doi.org/10.1016/S0065-3454(08)60380-0

Smith, J. E., & Holekamp, K. E. (2018). Spotted Hyenas. *Encyclopedia of Animal Behavior*, 3, 190–208. https://doi.org/10.1016/b978-0-12-809633-8.20749-8

Smith, J. E., Kolowski, J. M., Graham, K. E., Dawes, S. E., & Holekamp, K. E. (2008). Social and ecological determinants of fission–fusion dynamics in the spotted hyaena. *Animal Behaviour*, 76(3), 619–636. https://doi.org/10.1016/j.anbehav.2008.05.001

Smith, J. E., Van Horn, R. C., Powning, K. S., Cole, A. R., Graham, K. E., Memenis, S. K., & Holekamp, K. E. (2010). Evolutionary forces favoring intragroup coalitions among spotted hyenas and other animals. *Behavioral Ecology*, 21(2), 84–303.

Snell-Rood, E. C. (2013). An overview of the evolutionary causes and consequences of behavioural plasticity. *Animal Behaviour*, 85(5), 1004–1011. https://doi.org/10.1016/j.anbehav.2012.12.031

Snell-Rood, E. C., & Wick, N. (2013). Anthropogenic environments exert variable selection on cranial capacity in mammals. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769), 20131384. https://doi.org/10.1098/rspb.2013.1384

Sol, D. (2008). Do Successful Invaders Exist? Pre-Adaptations to Novel Environments in Terrestrial Vertebrates. In W. Nentwig (Ed.), *Biological Invasions. Ecological Studies (Analysis and Synthesis)* (vol 193, pp. 127–141). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-36920-2_8

Sol, D. (2009a). Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biology Letters*, 5(1), 130–133. https://doi.org/10.1098/rsbl.2008.0621

Sol, D. (2009b). The cognitive-buffer hypothesis for the evolution of large brains. In R. Dukas & J. M. Ratcliffe (Eds.), *Cognitive ecology II* (pp. 111–134). Chicago, IL: The University of Chicago Press.

Sol, D., Bacher, S., Reader, S. M., & Lefebvre, L. (2008). Brain Size Predicts the Success of Mammal Species Introduced into Novel Environments. *The American Naturalist*, 172(S1), S63–S71. https://doi.org/10.1086/588304

Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P., & Lefebvre, L. (2005). Big brains, enhanced cognition, and response of birds to novel environments. *Proceedings of the National Academy of Sciences of the United States of America*, 102(15), 5460–5465. https://doi.org/10.1073/pnas.0408145102

Sol, D., Lapiedra, O., & González-Lagos, C. (2013). Behavioural adjustments for a life in the city. *Animal Behaviour*, 85(5), 1101–1112. https://doi.org/10.1016/j.anbehav.2013.01.023

Sol, D., & Lefebvre, L. (2000). Behavioural flexibility predicts invasion success in birds introduced to New Zealand. *Oikos*, 90(3), 599–605. https://doi.org/10.1034/j.1600-0706.2000.900317.x

Sol, D., Sayol, F., Ducatez, S., & Lefebvre, L. (2016). The life-history basis of behavioural innovations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1690), 20150187-. https://doi.org/10.1098/rstb.2015.0187

Sol, D., Székely, T., Liker, A., & Lefebvre, L. (2007). Big-brained birds survive better in nature. *Proceedings of the Royal Society B: Biological Sciences*, 274(1611), 763–769. https://doi.org/10.1098/rspb.2006.3765

Sol, D., Timmermans, S., & Lefebvre, L. (2002). Behavioural flexibility and invasion success in birds. *Animal Behaviour*, 63(3), 495–502. https://doi.org/10.1006/anbe.2001.1953

Steele, T. E. (2013). VERTEBRATE RECORDS | Late Pleistocene of Africa. In S. A. Elias & C. J. Mock (Eds.), *Encyclopedia of Quaternary Science* (2nd edition, pp. 664–672). Elsevier. https://doi.org/10.1016/B978-0-444-53643-3.00247-8

Stoffel, M. A., Nakagawa, S., & Schielzeth, H. (2017). rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 8(11), 1639–1644. https://doi.org/10.1111/2041-210X.12797

Strauss, E. D., & Holekamp, K. E. (2019). Social alliances improve rank and fitness in convention-based societies. *Proceedings of the National Academy of Sciences*, 116(18), 8919–8924. https://doi.org/10.1073/pnas.1810384116

Swanson, E. M., Dworkin, I., & Holekamp, K. E. (2011). Lifetime selection on a hypoallometric size trait in the spotted hyena. *Proceedings of the Royal Society B: Biological Sciences*, 278(1722), 3277–3285. https://doi.org/10.1098/rspb.2010.2512

Swanson, E. M., Holekamp, K. E., Lundrigan, B. L., Arsznov, B. M., & Sakai, S. T. (2012). Multiple Determinants of Whole and Regional Brain Volume among Terrestrial Carnivorans. *PLoS One*, 7(6), e38447. http://dx.doi.org/10.1371/journal.pone.0038447

Tabachnick, B. G., & Fidell, L. S. (Eds.). (2013). Using multivariate statistics (6th ed.). Boston, MA: Pearson Education.

Tanner, J. B., Smale, L., & Holekamp, K. E. (2007). Ontogenetic variation in the play behavior of spotted hyenas. *Journal of Developmental Processes*, 2, 5–30.

Taylor, A. H., Hunt, G. R., Medina, F. S., & Gray, R. D. (2009). Do New Caledonian crows solve physical problems through causal reasoning? *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1655), 247–254. https://doi.org/10.1098/rspb.2008.1107

Therneau, T. M. (2015). A package for survival analysis in S. Retrieved from https://cran.r-project.org/package=surviva

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling survival data: Extending the cox model*. New York, NY: Springer.

Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: linking individual cognitive differences to fitness. *Behavioral Ecology*, 25(6), 1299–1301. https://doi.org/10.1093/beheco/aru095

Thornton, Alex, & Samson, J. (2012). Innovative problem solving in wild meerkats. *Animal Behaviour*, 83(6), 1459–1468. https://doi.org/10.1016/j.anbehav.2012.03.018

Titulaer, M., van Oers, K., & Naguib, M. (2012). Personality affects learning performance in difficult tasks in a sex-dependent way. *Animal Behaviour*, 83(3), 723–730. https://doi.org/10.1016/j.anbehav.2011.12.020

Turner, J. W., LaFleur, R. M., Richardson, A. T., & Holekamp, K. E. (2020). Risk-taking in free-living spotted hyenas is associated with anthropogenic disturbance, predicts survivorship, and is consistent across experimental contexts. *Ethology*, 126(1), 97–110. https://doi.org/10.1111/eth.12964

Turner, Julie Wolfe. (2018). Social development and its influence on adult traits in the spotted hyena (Doctoral dissertation, Michigan State University, East Lansing, USA). Available from ProQuest Dissertations & Theses Global database. (UMI No. 10809848).

van Horik, J. O., Langley, E. J. G., Whiteside, M. A., Laker, P. R., Beardsworth, C. E., & Madden, J. R. (2018). Do detour tasks provide accurate assays of inhibitory control? *Proceedings of the Royal Society B: Biological Sciences*, 285(1875). http://dx.doi.org/10.1098/rspb.2018.0150

Van Horik, J. O., Langley, E. J. G., Whiteside, M. A., Laker, P. R., & Madden, J. R. (2018). Intra-individual variation in performance on novel variants of similar tasks influences single factor explanations of general cognitive processes. *Royal Society Open Science*, 5(7). https://doi.org/10.1098/rsos.171919

van Horik, J. O., Langley, E. J. G., Whiteside, M. A., & Madden, J. R. (2017). Differential participation in cognitive tests is driven by personality, sex, body condition and experience. *Behavioural Processes*, 134, 22–30. https://doi.org/10.1016/j.beproc.2016.07.001

van Horik, J. O., & Madden, J. R. (2016). A problem with problem solving: motivational traits, but not cognition, predict success on novel operant foraging tasks. *Animal Behaviour*, 114, 189–198. http://dx.doi.org/10.1016/j.anbehav.2016.02.006

Van Horn, R. C., McElhinny, T. L., & Holekamp, K. E. (2003). Age Estimation and Dispersal in the Spotted Hyena (*Crocuta crocuta*). *Journal of Mammalogy*, 84(3), 1019–1030. https://doi.org/10.1644/BBa-023

Van Meter, P. E., French, J. a, Dloniak, S. M., Watts, H. E., Kolowski, J. M., & Holekamp, K. E. (2009). Fecal glucocorticoids reflect socio-ecological and anthropogenic stressors in the lives of wild spotted hyenas. *Hormones and Behavior*, 55(2), 329–337. https://doi.org/10.1016/j.yhbeh.2008.11.001

van Schaik, C. P., & Burkart, J. M. (2011). Social learning and evolution: the cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366(1567), 1008–1016. https://doi.org/10.1098/rstb.2010.0304

van Schaik, C. P., Isler, K., & Burkart, J. M. (2012). Explaining brain size variation: from social to cultural brain. *Trends in Cognitive Sciences*, 16(5), 277–284. https://doi.org/10.1016/j.tics.2012.04.004

van Woerden, J. T., van Schaik, C. P., & Isler, K. (2010). Effects of Seasonality on Brain Size Evolution: Evidence from Strepsirrhine Primates. *The American Naturalist*, 176(6), 758–767. https://doi.org/10.1086/657045

van Woerden, J. T., Willems, E. P., van Schaik, C. P., & Isler, K. (2012). Large brains buffer energetic effects of seasonal habitats in catarrhine primates. *Evolution*, 66(1), 191–199. https://doi.org/10.1111/j.1558-5646.2011.01434.x

Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression. *American Journal of Epidemiology*, 165(6), 710–718. http://dx.doi.org/10.1093/aje/kwk052

Völter, C. J., Tinklenberg, B., Call, J., & Seed, A. M. (2018). Comparative psychometrics: establishing what differs is central to understanding what evolves. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1756). http://dx.doi.org/10.1098/rstb.2017.0283

Vullioud, C., Davidian, E., Wachter, B., Rousset, F., Courtiol, A., & Höner, O. P. (2019). Social support drives female dominance in the spotted hyaena. *Nature Ecology & Evolution*, 3, 71–76. https://doi.org/10.1038/s41559-018-0718-9

Watts, H. E., & Holekamp, K. E. (2009). Ecological Determinants of Survival and Reproduction in the Spotted Hyena. *Journal of Mammalogy*, 90(2), 461–471. https://doi.org/10.1644/08-mamm-a-136.1

Webster, S. J., & Lefebvre, L. (2001). Problem solving and neophobia in a columbiform–passeriform assemblage in Barbados. *Animal Behaviour*, 62(1), 23–32. https://doi.org/10.1006/anbe.2000.1725

Weisbecker, V., Blomberg, S., Goldizen, A. W., Brown, M., & Fisher, D. (2015). The Evolution of Relative Brain Size in Marsupials Is Energetically Constrained but Not Driven by Behavioral Complexity. *Brain, Behavior and Evolution*, 85(2), 125–135. https://doi.org/10.1159/000377666

Wetzel, D. P. (2017). Problem-solving skills are linked to parental care and offspring survival in wild house sparrows. *Ethology*, 123(6–7), 475–483. https://doi.org/10.1111/eth.12618

Whiten, A. (2018). Social, Machiavellian and cultural cognition: A golden age of discovery in comparative and evolutionary psychology. *Journal of Comparative Psychology*, 132(4), 437.

Whiten, A., & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences*, 11(2), 233–273. https://doi.org/10.1017/S0140525X00049682

Whitley, E., Deary, I. J., Ritchie, S. J., Batty, G. D., Kumari, M., & Benzeval, M. (2016). Variations in cognitive abilities across the life course: Cross-sectional evidence from Understanding Society: The UK Household Longitudinal Study. *Intelligence*, 59, 39–50. https://doi.org/10.1016/j.intell.2016.07.001

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, 75(5), 1182–1189. https://doi.org/10.1111/j.1365-2656.2006.01141.x

Willemet, R. (2013). Reconsidering the evolution of brain, cognition, and behavior in birds and mammals. *Frontiers in Psychology*, 4(396), 10–3389. https://doi.org/10.3389/fpsyg.2013.00396

Wisnieski, L., Norby, B., Pierce, S. J., Becker, T., Gandy, J. C., & Sordillo, L. M. (2019). Predictive models for early lactation diseases in transition dairy cattle at dry-off. *Preventive Veterinary Medicine*, 163, 68–78. https://doi.org/https://doi.org/10.1016/j.prevetmed.2018.12.014

Wolf, M., van Doorn, G. S., Leimar, O., & Weissing, F. J. (2007). Life-history trade-offs favour the evolution of animal personalities. *Nature*, 447(7144), 581–584. https://doi.org/10.1038/nature05835

Woodley of Menie, M. A., Fernandes, H. B. F., & Peñaherrera Aguirre, M. (2017). General Intelligence Factor G (Reader, Hager, and Laland, 2011). In T. K. Shackelford & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of Evolutionary Psychological Science* (pp. 1–4). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-16999-6_3096-1

Wright, T. F., Eberhard, J. R., Hobson, E. A., Avery, M. L., & Russello, M. A. (2010). Behavioral flexibility and species invasions: the adaptive flexibility hypothesis. *Ethology Ecology & Evolution*, 22(4), 393–404. https://doi.org/10.1080/03949370.2010.505580

Yirga Abay, G., Bauer, H., Gebrihiwot, K., & Deckers, J. (2010). Peri-urban spotted hyena (Crocuta crocuta) in Northern Ethiopia: diet, economic impact, and abundance. *European Journal of Wildlife Research*, 57(4), 759–765. https://doi.org/10.1007/s10344-010-0484-8

Yirga, G., De Iongh, H. H., Leirs, H., Gebrehiwot, K., Berhe, G., Asmelash, T., … Bauer, H. (2013). The ecology of large carnivores in the highlands of northern Ethiopia. *African Journal of Ecology*, 51(1), 78–86. https://doi.org/10.1111/aje.12008

Yirga, G., Ersino, W., De Iongh, H. H., Leirs, H., Gebrehiwot, K., Deckers, J., & Bauer, H. (2013). Spotted hyena (Crocuta crocuta) coexisting at high density with people in Wukro district, northern Ethiopia. *Mammalian Biology - Zeitschrift Für Säugetierkunde*, 78(3), 193–197. https://doi.org/10.1016/j.mambio.2012.09.001

Yirga, G., Leirs, H., De Iongh, H. H., Asmelash, T., Gebrehiwot, K., Vos, M., & Bauer, H. (2017). Densities of spotted hyaena (Crocuta crocuta) and African golden wolf (Canis anthus) increase with increasing anthropogenic influence. *Mammalian Biology*, 85, 60–69. https://doi.org/10.1016/j.mambio.2017.02.004

Yoerg, S. I. (1991). Social feeding reverses learned flavor aversions in spotted hyenas (Crocuta crocuta). *Journal of Comparative Psychology*, 105(2), 185–189. https://doi.org/10.1037/0735-7036.105.2.185

Yoshida, K. C. S., Van Meter, P. E., & Holekamp, K. E. (2016). Variation among free-living spotted hyenas in three personality traits. *Behaviour*, 153(13–14), 1665–1722. https://doi.org/10.1163/1568539X-00003367

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x