Cognition in the wild

Individual differences in cognitive abilities and their link with fitness in a wild primate, the grey mouse lemur (*Microcebus murinus*)

Dissertation

for the award of the degree
"Doctor rerum naturalium" (Dr.rer.nat.)
of the Georg-August-Universität Göttingen

within the doctoral program Behavior and Cognition (BeCog)
of the Georg-August University School of Science (GAUSS)

submitted by
Franziska Hübner
from Munich

Göttingen, 2019

Thesis Committee

Dr. Claudia Fichtel, Behavioral Ecology and Sociobiology, German Primate Center

Prof. Dr. Lars Penke, Biological Personality Psychology, University of Göttingen

Dr. Oliver Schülke, Behavioral Ecology, University of Göttingen

Members of the Examination Board

First Reviewer:

Dr. Claudia Fichtel, Behavioral Ecology and Sociobiology, German Primate Center

Second Reviewer:

Prof. Dr. Lars Penke, Biological Personality Psychology, University of Göttingen

Further members of the Examination Board

Dr. Oliver Schülke, Behavioral Ecology, University of Göttingen

Prof. Dr. Peter Kappeler, Sociobiology/Anthropology, University of Göttingen, and Behavioral Ecology and Sociobiology, German Primate Center

Prof. Dr. Julia Fischer, Cognitive Ethology, German Primate Center

Prof. Dr. Margarete Boos, Social and Communication Psychology, University of Göttingen

Date of oral examination: 23rd of May 2019

Contents

Summary

Cognition, the process by which animals acquire, process, store and use information from their environment, plays a major role in various behaviours across all aspects of an individuals' life. However, despite this overall importance of cognitive abilities, we still know little about how cognitive traits evolved. In the past, research on cognitive evolution focussed on comparing different species in their cognitive abilities and linked variation in cognition to inter-specific differences in ecological and social conditions. This comparative approach helps to understand when in evolutionary history and under which conditions particular cognitive abilities evolved. Individual variation in cognitive abilities has been widely ignored in these studies on the species level, however. This changed only recently when the interest in how and why individuals differ in cognitive traits emerged. Only by investigating inter-individual variation in cognitive abilities and their link with fitness outcomes, we can begin to understand the causes and consequences of this variation and finally unravel how cognition evolved. Because fitness can only be studied in wild, free-ranging individuals, the study of individual differences in cognitive abilities and their fitness consequences is challenging and has been conducted mainly in different species of birds until today. The few studies so far focused largely on single measures of cognition and fitness and revealed not only positive but also negative and not significant correlations between cognitive performance and fitness outcomes.

With my thesis, I contribute to this young field of research aiming to better understand the adaptive value of cognitive traits. I investigated cognitive abilities in five cognitive tasks in wild grey mouse lemurs (*Microcebus murinus*), a primate species endemic to Madagascar, and linked cognitive performance with different fitness proxies. The cognitive tasks addressed different ecologically relevant cognitive abilities. They included associative and motor learning during repeated problem solving, causal reasoning and spontaneous problem solving in a string-pulling task, spatial learning, inhibitory control and behavioural flexibility during reversal learning. Subjects were tested during short-term captivity of up to three nights and subsequently released back into the wild. Furthermore, individuals were tested in two personality tests and their neophilia and activity during a novel object and an open field test was assessed.

In the first part of my thesis, I focused on inter- and intra-individual differences of cognitive performance: the potential confounding effects of non-cognitive factors on cognitive performance and the structure of cognitive performance across different cognitive abilities. My results demonstrated that, first of all, subjects differed greatly in cognitive performance and performance was not systematically affected by non-cognitive factors such as personality, body condition, motivation, age or sex. Furthermore, performance in one cognitive task was generally a weak predictor of performance in any other tasks of the test battery and could not be summarized into a general factor. Thus, I could not find evidence for a general factor of cognitive performance similar to the general intelligence factor (g) in humans, where performance is positively correlated across cognitive tasks and domains.

In the second part of the thesis, I studied different fitness-related traits in grey mouse lemurs and linked cognitive performance with two fitness proxies. I found that efficient repeated solving of a food extraction task correlated positively with the ability to maintain body condition during the long dry season with low food availability. This suggests that the ability to quickly apply a newly discovered motor technique during problem solving also facilitates the exploitation of new, natural food resources when food is scarce. By contrast, cognitive performance, irrespective of cognitive task, did not correlate with survival or longevity, suggesting that the assessed cognitive abilities did not provide (net) benefits in survival. Moreover, difficulties in validating a physiological condition factor, that summarized measures of body condition, hematocrit and long-term cortisol levels, demonstrated that identifying and operationalizing meaningful fitness proxies can be challenging for many taxa.

In summary, my thesis contributes the first study on a wild primate to the growing body of research investigating individual differences in cognitive abilities and their link with fitness outcomes. Together with the other recent findings revealing heterogenous links between cognitive performance and fitness measures, my thesis demonstrates that cognitive abilities are involved in complex interactions between various traits. At the same time, they are likely to have costs and benefits, and thus do not necessarily correlate positively with fitness outcomes. Further studies in different species in the wild, that investigate multiple cognitive traits and fitness outcomes as well as potentially confounding covariates simultaneously, will help to unravel this complex system, the evolution of cognition.

Zusammenfassung

Kognition, der Prozess durch den Tiere Informationen aus ihrer Umwelt erhalten, verarbeiten, speichern und nutzen, spielt eine existenzielle Rolle in unterschiedlichsten Verhaltensweisen in allen Lebensbereichen eines Individuums. Trotz dieser umfassenden Bedeutung kognitiver Fähigkeiten wissen wir jedoch bisher immer noch wenig darüber, wie Kognition im Laufe der Evolution entstanden ist. In der Vergangenheit hat sich die Forschung im Bereich der Evolution von Kognition darauf fokussiert, verschiedene Arten in ihren kognitiven Fähigkeiten zu vergleichen und diese Unterschiede mit Gegensätzen im ökologischen und sozialen Umfeld der Arten zu erklären. Dieser vergleichende Ansatz hilft zu verstehen, wann in der Geschichte der Evolution und unter welchen Bedingungen bestimmte kognitive Fähigkeiten entstanden sind. Individuelle Unterschiede in kognitiven Fähigkeiten wurden in diesen Studien mit Fokus auf den Unterschieden zwischen Arten jedoch weitestgehend ignoriert. Dies änderte sich erst vor kurzem, als das Interesse dafür, wie und warum Individuen sich in ihren kognitiven Fähigkeiten unterscheiden, geweckt wurde. Nur wenn wir Kognitionsunterschiede zwischen einzelnen Individuen untersuchen und mit Fitnessmerkmalen in Verbindung setzen, können wir beginnen die Ursachen und Konsequenzen von kognitiven Fähigkeiten zu verstehen und herausfinden wie Kognition entstanden ist. Da Fitness nur bei freilebenden Tieren realistisch zu messen ist, geht dieses Vorhaben mit einigen Herausforderungen einher und wurde bis heute vor allem mit verschiedenen Vogelarten durchgeführt. Die wenigen Studien bisher haben sich vorwiegend auf einzelne Kognitions- und Fitnessmaße konzentriert und brachten positive, negative und nicht signifikante Korrelationen zwischen kognitiven Fähigkeiten und Fitnessmerkmalen zum Vorschein.

Mit meiner Doktorarbeit trage ich zu diesem jungen Wissenschaftsbereich bei, mit dem Ziel den adaptiven Wert kognitiver Fähigkeiten besser zu verstehen. Ich habe kognitive Fähigkeiten bei wildlebenden Grauen Mausmakis (*Microcebus murinus*), einer Affenart aus Madagaskar, untersucht und mit verschieden Fitnessmaßen korreliert. In fünf verschiedenen Kognitionstests wurden unterschiedliche, ökologisch relevante kognitive Fähigkeiten getestet. Sie umfassten Assoziationslernen und Bewegungslernen während wiederholtem Problem-lösen, kausales Verständnis und spontanes Problemlösen während eines String-Pulling Tests, räumliches Lernen, inhibitorische Kontrolle und flexibles Verhalten innerhalb eines Reversal Learning Tests. Die Versuchstiere wurden in Kurzzeit-Gefangenschaft von bis zu drei Nächten

getestet und im Anschluss wieder in ihren Streifgebieten freigelassen. Des Weiteren wurden die Tiere in zwei Persönlichkeitstest getestet und ihre Neophilie und Aktivität in einem Novel Object und einem Open Field Test gemessen.

Im ersten Teil meiner Arbeit ging es im Detail um die individuelle Variation in kognitiven Fähigkeiten und den potentiellen Einfluss von nicht kognitiven Faktoren auf die Performanz der Tiere. Außerdem ging es um darum, wie sich die kognitiven Fähigkeiten der Tiere in den verschiedenen Tests unterscheiden, der so genannten Struktur von Kognition. Meine Ergebnisse zeigen, dass die Tiere deutlich in ihren kognitiven Fähigkeiten variieren und kognitive Performanz nicht systematisch durch Persönlichkeitsmerkmale, Körperkondition, Motivation, Geschlecht und Alter beeinflusst wurde. Wie ein Tier in einem speziellen Test abschnitt, sagte nicht vorher wie es in den jeweils anderen Tests abschnitt und die Performanz der verschiedenen Tests konnte nicht zu einem generellen Faktor zusammengefasst werden. Daher konnte ich keinen Nachweis für einen generellen Faktor entsprechend dem General Intelligence Factor (g) beim Menschen, wo kognitive Fähigkeiten verschiedener Tests und kognitiven Domänen korrelieren, finden.

Im zweiten Teil meiner Arbeit habe ich verschiedene Fitness-relevante Merkmale bei Grauen Mausmakis untersucht und die kognitiven Fähigkeiten der Tiere mit zwei Fitnessmaßen in Verbindung gesetzt. Ich fand heraus, dass effizientes, wiederholtes Lösen eines Food Extraction Tasks signifikant mit der Fähigkeit die Körperkondition während der nahrungsarmen Trockenzeit aufrechtzuerhalten korrelierte. Dies deutet darauf hin, dass die Fähigkeit schnell eine neuentdeckte motorische Technik anzuwenden auch unter natürlichen Bedingungen von Vorteil ist und möglicherweise beim Erschließen neuer, natürlicher Nahrungsquellen eine Rolle spielt. Im Gegensatz dazu haben die kognitiven Fähigkeiten in den durchgeführten Tests nicht mit dem Überleben oder der Lebensdauer der Tiere korreliert, was darauf hindeutet, dass die erfassten kognitiven Fähigkeiten keinen (Gesamt-)Vorteil beim Überleben mit sich bringen. Außerdem zeigten Schwierigkeiten beim Validieren eines Maßes der physiologischen Verfassung der Tiere, welches körperliche Kondition, Hämatokrit und Langzeit-Kortisolwerte zusammenfasste, dass es eine Herausforderung sein kann für manche Taxa geeignete und bedeutsame Fitnessmaße zu finden.

Meine Arbeit ist die erste Studie, die individuelle Variation in kognitiven Fähigkeiten und deren Zusammenhang mit Fitnessmerkmalen bei einer wildlebenden Affenart untersucht hat und trägt damit zur wachsenden Forschung bei. Zusammen mit den jüngsten Ergebnissen anderer Studien zeigen meine Ergebnisse, dass kognitive Fähigkeiten in vielschichtigen, komplexen Interaktionen mit verschieden Merkmalen eines Tieres agieren. Gleichzeitig haben bessere kognitive Fähigkeiten sehr wahrscheinlich nicht nur Vorteile sondern auch Nachteile, was erklärt, warum sie nicht notwendigerweise positiv mit Fitnessmerkmalen korrelieren. Um diese komplexen Zusammenhänge in ihrer Gänze zu verstehen, werden wir noch weitere Studien benötigen, die gleichzeitig unterschiedliche kognitive Fähigkeiten, Fitnessmaße und andere, möglicherweise korrelierende Merkmale in verschiedenen Arten untersuchen. Dies wird letztendlich helfen die Evolution von Kognition zu verstehen.

General Introduction

Belonging to the apparently most intelligent, dominant and influential species on this planet, probably no human being would doubt that being smart pays off and made us hold this position. We possess one of the largest brains relative to body mass (Roth & Dicke, 2005), are superior innovators and problem-solvers, create art, use language and teaching, our knowledge accumulates over generations and we cooperate across the world (Shettleworth, 2012; MacLean, 2016). In order to understand what sets us apart and how our cognitive capacities evolved, comparing humans' and nonhuman animals' cognitive abilities has been the focus of cognitive research for the last decades (e.g. Herrmann et al., 2007; MacLean et al., 2012).

Yet, when looking around, it becomes clear that there are individual differences in people's cognitive capacities and not all individuals are able to learn or solve problems to the same extent. Indeed, individual differences in human cognitive abilities have been extensively studied: intelligence is heritable and stable across developmental stages as well as long timespans (Deary et al., 2010). Moreover, inter-individual variation in general intelligence is a good predictor of important life outcomes, like school and occupational achievement, social mobility, health and survival (Deary et al., 2010; Plomin & Deary, 2015).

Cognition is the neuronal process with which individuals acquire, process, store and use information from their environment (Shettleworth, 2010). Consequently, cognitive processes are involved in various behaviours across different contexts ranging from foraging to predator avoidance, but yet, the evolution of cognition remains poorly understood. Also in animals, stable individual differences in cognitive abilities exist (Cauchoix et al., 2018). However, research has only recently begun to pay attention to this inter-individual variation that provides the material on which selection can act (Thornton & Lukas, 2012) and the question if it actually pays off to be smart has rarely been asked. Today, few studies exist that linked individual variation in cognitive performance with fitness outcomes and also the not to be underestimated pitfalls and challenges that go along with this undertaking have been addressed (Rowe & Healy, 2014; Thornton et al., 2014; Morand-Ferron et al., 2016; Boogert et al., 2018). Nonetheless, only by acknowledging individual variation in cognition and investigating its fitness consequences, we can start to understand how selection acts on cognitive traits and how cognition evolved.

With my thesis, I contribute to this endeavour by investigating individual variation in cognitive abilities in a primate species and by linking cognitive performance with different fitness proxies in the wild. To begin with, in this introduction, I will start to review recent research approaches and insights into the evolution of cognition. I will focus on the importance of inter- and intra-individual variation in cognitive abilities for the study of cognitive evolution and address the fitness consequences of individual differences in cognition. Further, I will consider the challenges in quantifying individual cognitive variation and in studying the cognition-fitness link. Subsequently, I will introduce the study species, the grey mouse lemur (*Microcebus murinus*) and illustrate its suitability for studying the link between cognition and fitness in a wild primate. Finally, I will outline the specific aims of this thesis.

Understanding the evolution of cognition

From comparative cognition to individual differences

Traditionally, research in cognition focused on understanding proximate mechanisms underpinning cognitive processes. Universal learning principles have been elucidated by studying few model species (e.g. rats and pigeons) that could be trained in elaborate cognitive tasks (Skinner, 1938). Furthermore, the adaptive value of cognition has been addressed by comparing species in their cognitive abilities or proxies for cognitive capacity, such as measures of brain size, and linking this interspecific variation to differences in ecological and social factors (Bshary et al., 2002; Lefebvre et al., 2004; Dunbar & Shultz, 2007; Maclean et al., 2008; Scheid & Bugnyar, 2008; Dechmann & Safi, 2009; Gonzalez-Voyer et al., 2009; MacLean et al., 2014). This comparative approach helped to understand when in evolutionary history and under which conditions particular cognitive abilities evolved. Several, mutually non-exclusive hypotheses have been discussed and empirically supported. For example, the ecological intelligence hypothesis predicts that dietary complexity and reliance on spatially dispersed fruits shaped primate cognitive evolution and correlates with interspecific differences in cognitive abilities (Clutton-Brock & Harvey, 1980; Milton, 1981), but also finds support in other taxa when linking feeding ecology with cognition (Balda & Kamil, 1989; Shettleworth, 1990; Barkley & Jacobs, 2007). In contrast, the social intelligence hypothesis proposes that differences in social complexity predict interspecific variation in cognitive abilities and brain sizes, resembling another major driver of primate cognitive evolution (Humphrey, 1976; Byrne & Whiten, 1988; Dunbar, 1998). Evidence for convergent and divergent evolutionary processes have been provided (Emery & Clayton, 2004; Healy et al., 2005; Lefebvre & Sol, 2008; Reader et al., 2011; MacLean et al., 2012; Van Horik et al., 2012) as well as for rapid cognitive divergence between populations of the same species facing different selective pressures in the wild (Pravosudov & Clayton, 2002; Brown & Braithwaite, 2005; Roth et al., 2010; Roth et al., 2012).

Typically, when comparing different species in their cognitive abilities, only a few captive individuals are tested in the same or similar cognitive tasks and sometimes even highly enculturated “genius” individuals serve to represent their whole species’ cognitive capacities (Thornton & Lukas, 2012; Boogert et al., 2018). With this emphasis on species-specific cognitive abilities, individual variation is treated as noise around the population mean, resulting in a binary perspective of the presence or absence of a given cognitive ability (Thornton & Samson, 2012). However, also within a given species, individuals differ remarkably in their cognitive abilities and only by acknowledging this inter-and intra-individual variation, we can really begin to understand the evolution of cognition through natural selection (Thornton & Lukas, 2012; Morand-Ferron et al., 2016; Boogert et al., 2018).

Individual variation in cognitive abilities

Studies in captivity reveal strong evidence that individuals differ in their cognitive abilities (Reader, 2003; Dukas, 2004; Boogert, et al., 2011; Thornton & Lukas, 2012). Since Darwin recognized that “no one supposes that all the individuals of the same species are cast in the same actual mould”, we know that individual differences are of critical importance in evolution as they “afford materials for natural selection to act on” (Darwin, 1859, pp. 59-60). Thus, variation in cognitive traits constitutes the first of three necessary conditions for selection and evolution of cognition (Darwin, 1859, 1871).

Second, as with any other phenotypic trait, selection can only act on cognition if cognitive traits are heritable and if there is an underlying genetic variation. Substantial heritability of cognitive traits has been demonstrated for example in artificial selection experiments with model species (Wahlsten, 1972; Plomin, 2001; Dukas, 2004, 2008; Kawecki, 2010; Mery, 2013; Croston et al., 2015; Smith et al., 2015). Moreover, in humans, more than half of the individual differences in intelligence have been attributed to additive genetic variation (Deary et al., 2010; Davies et al., 2011). However, for most taxa and especially for wild populations, estimates of heritability are still lacking (Thornton et al., 2014).

Third, variation in cognitive traits must result in fitness differences. An adaptive value of cognitive abilities has been suggested in various different contexts as cognitive abilities drive a wide range of behaviours with various fitness consequences, ranging from reproduction (Hollis et al., 1989; Boogert, et al., 2011; Minter et al., 2017), predator avoidance (Griffin, 2004; Lonnstedt et al., 2012), social interactions (Domjan et al., 2000; Hansen & Slagsvold, 2004), navigation (Dyer, 1998) to foraging (Slagsvold & Wiebe, 2007; Rahmani et al., 2009; Morand-Ferron, 2017). In conclusion, the finding that there is heritable variation in cognitive abilities between individuals which is likely to be related to variance in fitness indicates that cognitive abilities can be acted upon by natural and sexual selection. However, direct evidence for the selection of cognitive abilities in natural populations is still lacking (Morand-Ferron et al., 2016).

Links between cognitive ability and fitness

A handful of studies started to link variation in cognitive performance with proxies of fitness in wild animals of different species to examine how selection might act on cognitive abilities (for an overview see also Table S1 in Chapter 2, supplementary material). A positive correlation between learning speed and colonies' overall foraging success was found in bumblebee colonies (*Bombus terrestris*) (Raine & Chittka, 2008) but individual bumblebees' learning ability did not correlate with daily foraging performance in another study (Evans et al., 2017). In captive male bitterlings (*Rhodeus ocellatus*), spatial learning accuracy correlated positively with reproductive success in a sneaker role, but not in the dominant guarding role, the alternative male mating tactic in this fish species (Smith et al., 2015).

Problem-solving performance was used as a cognitive performance measure in studies with various bird species in the wild, presenting subjects with novel problems, like artificial foraging tasks. In great tits (*Parus major*) (Cole et al., 2012; Cauchard et al., 2013; Preiszner et al., 2017) and house sparrows (*Passer domesticus*) (Wetzel, 2017), problem-solving performance or success correlated positively with measures of reproductive success, but problem-solver pairs were more likely to desert their nest, resembling a fitness cost associated with better cognitive performance (Cole et al., 2012). Moreover, problem-solving performance did not correlate with survival of adults (Cole et al., 2012). Also in Australian magpies (*Cracticus tibicen dorsalis*), general cognitive performance in four different tasks predicted reproductive success in females (Ashton et al., 2018). In bowerbirds, males' problem-solving performance in one species was positively correlated with mating success (*Ptilonorhynchus violaceus*) (Keagy et al., 2009, 2012), while cognitive performance in six different tests did not correlate with males' mating success in another closely related species (*Ptilonorhynchus maculatus*) (Isden et al., 2013). Furthermore, in song sparrows (*Melospiza melodia*), reversal learning performance correlated positively with male song repertoire size, a predictor of various fitness-related traits, whereas motor and associative learning performance did not, and detour-reaching performance was negatively correlated to song repertoire size (Boogert et al., 2011). In contrast, in common pheasants (*Phasianus colchicus*) reversal learning performance was negatively correlated with fitness. Individuals that were slow to reverse a learned association were more likely to survive for 60 days under semi-wild conditions (Madden et al., 2018). In addition, links between associative learning and survival probability depended on the weight of birds; heavy pheasants that were quick in learning associations were more likely to survive, whereas for light individuals, quick associative learners were less likely to survive for 60 days (Madden et al., 2018). Finally, for female African striped mice (*Rhabdomys pumilio*), short-term spatial memory performance correlated negatively with survival. In contrast, in males, better long-term spatial memory performance correlated positively with the number of days subjects survived until the breeding season (Maille & Schradin, 2016).

Taken together, the strengths and directions of correlations between cognitive performance and fitness proxies differ between studies and cannot be generalized across cognitive and fitness measures, but also not across and within species. This may partially be explained by differences in study design as studies vary in regard to the investigated cognitive mechanisms and tasks applied but also in the fitness proxies assessed. So far, most studies were conducted with different species of birds and only one study investigated fitness correlates of cognitive performance in a mammal (Maille & Schradin, 2016). Moreover, the majority of studies focussed on testing subjects (often of one sex) in only one cognitive ability and linked performance with variation in a single fitness measure. Thus, current findings indicate that drawing general conclusions on the fitness consequence of cognitive abilities is still difficult but may also not be possible regarding the complex interactions and trade-offs cognitive abilities are involved in, which will be explained in the following.

Importantly, not only benefits are associated with better cognitive abilities but higher cognitive performance is also correlated with costs resulting in fitness trade-offs. Especially the development and maintenance of neuronal structures is energetically very costly (Kawecki, 2010; Kotrschal et al., 2013), but also information gathering and processing costs time and energy (Dukas & Visscher, 1994; Laughlin et al., 1998; Laughlin, 2001; Jaumann et al., 2013; Evans et al., 2017). Artificial selection experiments with fruit flies (*Drosophila melanogaster*) and mice (*Mus musculus*), but also studies with wild populations detected fitness disadvantages that correlated with better cognitive performance, like reduction in immunity (Barnard et al., 2006), reduced longevity (Burger et al., 2008) and lower reproductive success (Mery & Kawecki, 2003; Snell-Rood et al., 2011).

Cognitive traits might also be correlated with other behavioural traits, such as personality traits (reviewed in Øverli et al., 2007; Carere & Locurto, 2011), which themselves are likely to impact fitness and could mask links between cognition and fitness (see Fig. 1) (Morand-Ferron & Quinn, 2015; Morand-Ferron et al., 2016). Thus, associated fitness trade-offs and correlated traits demonstrate that better cognitive abilities will only be selected if they result in net benefits of fitness. Moreover, cognition is not a unitary trait but any given behaviour requires multiple cognitive processes and also a specific cognitive ability is involved in various different contexts (see Fig. 1) (Rowe & Healy, 2014). What is beneficial in one situation, might not be beneficial in another as it depends on the time, context and environment (Rowe & Healy, 2014; Ten Cate, 2014). Ultimately, these costs and benefits of cognition, as well as the complex relationships between various different traits may explain why individual variation in cognition is maintained (Morand-Ferron & Quinn, 2015). Furthermore, they indicate that cognitive performance in a particular test may not necessarily be closely and positively correlated with a given fitness measure, and detecting the underlying trade-offs can be especially challenging in the wild (Rowe & Healy, 2014; Morand-Ferron et al., 2016; but see Cole et al., 2012).

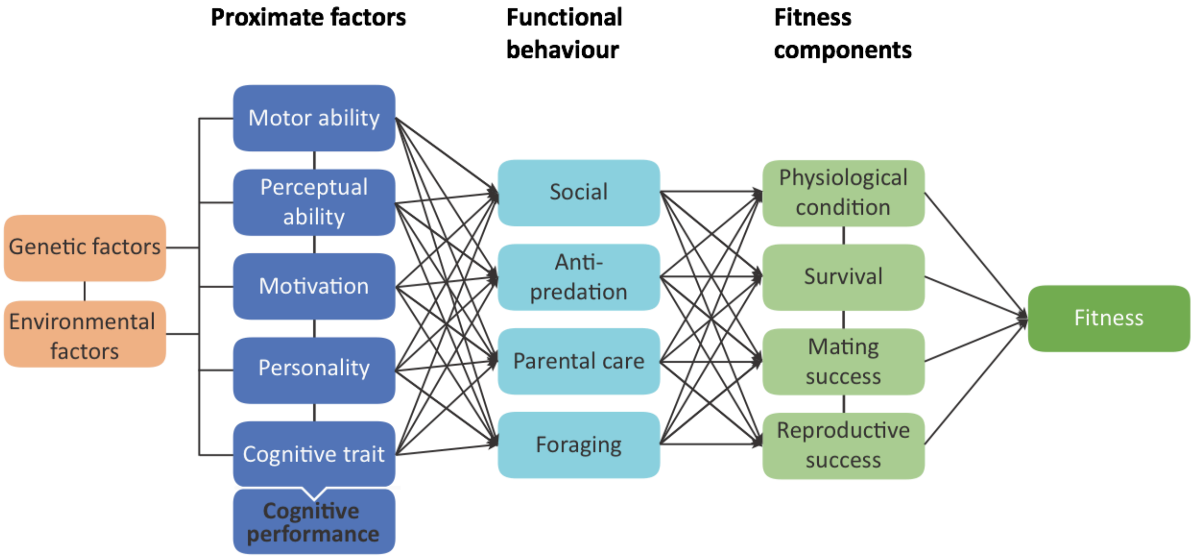

Figure 1: The potential complex pathways that link cognition with fitness. Pathways from genetic and environmental factors to cognitive abilities and other proximate factors, impacting behaviours that influence different fitness components which result in the final fitness of an individual; its genetic contribution to future generations. Importantly, proximate factors are likely to interact with each other but also to influence cognitive performance in a cognitive task, the measure used to infer cognitive traits or abilities. The degree to which a given cognitive ability influences various functional behaviours may vary. For example in a food-caching species, individual variation in spatial memory could impact foraging during the food-scarce winter, which influences physiological condition and over-winter survival (e.g. Rowe & Healy, 2014). By contrast, individual variation in other cognitive abilities, such as associative learning, may be more universally influencing behaviours involved in various fitness aspects. Depicted pathways and lists of different elements are non-exhaustive. Adapted from Morand-Ferron et al., 2016.

Intra-individual variation in cognitive abilities

Another interesting aspect in understanding the evolution of cognition is how cognitive abilities are structured, i.e. correlated within individuals, and how these structures evolved. Humans possess domain-general intelligence which means that cognitive performance is positively correlated across tests assessing abilities in different cognitive domains such as reasoning, processing speed, executive function, memory and spatial ability (Deary et al., 2010; Burkart et al., 2017). In other words, individuals that perform well in one test are also good in other psychometric tests tapping into other cognitive domains (Deary et al., 2010). Statistically, these positive correlations between test performances can be extracted and reveal a single factor, the general intelligence factor (g), that explains a significant amount of variation in cognitive performance (Burkart et al., 2017).

By contrast, in nonhuman animals, evidence for general intelligence is still rare. A modular organisation with domain-specific adaptations to specific ecological problems has been emphasized in the past and might be the ancestral state of vertebrate cognition (Shettleworth, 2012; Burkart et al., 2017). For example, food caching species would evolve enhanced spatial memory abilities as a specific cognitive adaptation to the challenges of food cache recovery (Pravosudov & Roth, 2013). Studies across a handful of animal species, applying different cognitive tests to captive but also to few wild individuals, revealed mixed evidence for domain-general intelligence and the presence of a general intelligence factor analogous to human g (reviewed in Shaw & Schmelz, 2017). Evidence for a g factor has been reported for laboratory rodents (Matzel et al., 2003; Galsworthy et al., 2005; Kolata et al., 2008; Light et al., 2010; Wass et al., 2012; but see Locurto et al., 2003), dogs (Arden & Adams, 2016), and in some studies with different species of birds tested in the wild (Isden et al., 2013; Shaw et al., 2015; Ashton et al., 2018) but not in others (Boogert et al., 2011; Keagy et al., 2011; van Horik et al., 2018).

Especially for nonhuman primates, tested with large test batteries in the laboratory, evidence of g is mixed. A modular organization of the primate brain with domain-specific cognitive skills has been discussed (reviewed in Amici et al., 2012) but general intelligence might coexist with these domain-specific adaptations (Amici et al., 2017; Burkart et al., 2017). A g factor has been reported for cotton-top tamarins (*Saguinus Oedipus*) (Banerjee et al., 2009), orangutans (*Pongo pygmaeus wurmbii* and *Pongo abelii*) (Damerius et al., 2018), and in one study with chimpanzees (*Pan troglodytes*) (Hopkins et al., 2014). In contrast, two other studies on chimpanzees could not summarize individuals' performance across tasks and domains into a single g factor but found "clusters of cognitive abilities", thus evidence for domain-specific cognitive abilities (Herrmann et al., 2010; Herrmann & Call, 2012).

To summarize, until today evidence on g is mixed and drawing general conclusions across species and studies is still difficult. Comparisons across studies are also complicated because studies differ widely in the applied cognitive tasks and the number and kind of addressed cognitive domains. Moreover, motivation and other confounding factors that might systematically affect cognitive performance across tasks are rarely controlled for and different statistical methods were applied when drawing conclusions on g (Burkart et al., 2017; Shaw & Schmelz, 2017; van Horik et al., 2018; Völter et al., 2018). Thus, more studies are needed that also account for systematic effects of non-cognitive factors. Especially studies with wild, unexperienced subjects are promising in this regard as captive, human-reared individuals are often highly enculturated and have a long testing history which may bias general test performance (Thornton & Lukas, 2012). Once this additional data is gained, a more complete and coherent picture will advance our understanding how general intelligence evolved.

Quantifying individual variation in cognition

Importantly, cognitive abilities have to be reliably quantified as they cannot be observed directly but must be inferred through behaviour, i.e. performance in cognitive experiments (Shettleworth, 2010). Cognitive tasks should be designed to target defined cognitive processes and variation in cognitive performance due to confounding factors as well as random noise must be excluded (Rowe & Healy, 2014; Thornton et al., 2014; Morand-Ferron et al., 2016; Boogert et al., 2018).

Various non-cognitive factors, like rearing condition, prior experiences, motivation, persistence and personality can potentially affect how an individual performs in a cognitive test (Thornton & Lukas, 2012; Rowe & Healy, 2014; Morand-Ferron et al., 2016; Dougherty & Guillette, 2018). Especially motivation is likely to determine participation and performance in a cognitive task and especially in food motivated tasks subjects' feeding motivation might differ (Rowe & Healy, 2014; Morand-Ferron et al., 2016). Personality differences, i.e. stable, consistent individual variation in behaviour (Dall et al., 2004), can co-vary with the way animals acquire, process and store information, thus affecting cognitive processes directly (Carere & Locurto, 2011; Griffin et al., 2015; Sih & Del Giudice, 2012). For example, bold, fast exploring individuals are predicted to be quicker but less accurate in learning new contingencies than neophobic, less explorative individuals (Sih & Del Giudice, 2012). Furthermore, personality could affect cognitive performance by increasing probabilities or rates of exposure with the task and the learning contingencies (Carere & Locurto, 2011; Morand-Ferron et al., 2016; Guillette et al., 2017). Because it is difficult to exclude these confounding factors in most testing regimes, especially when testing wild individuals, assessing these differences in non-cognitive factors and controlling statistically for them is important to reliably quantify inter-individual variation in cognitive performance (Griffin & Guez, 2014; Morand-Ferron et al., 2016).

Testing performance in cognitive tasks repeatedly allows to demonstrate that individual variation in cognition is consistent across time and context and was not, for instance, influenced by temporary distraction (Boogert et al., 2018). Repeatability of cognitive performance across species and cognitive measures was shown to be low to moderate, either when the same task was presented twice (temporal repeatability), or when performance in different tasks addressing the same cognitive abilities was compared (contextual repeatability) (Cauchoix et al., 2018).

In order to detect links between cognitive abilities and fitness outcomes, first of all, studying ecologically relevant cognitive abilities, i.e. choosing cognitive abilities with respect to a species' natural history is important (Morand-Ferron et al., 2016). This means that targeted cognitive traits should be chosen with regard to a species' ecological problems and the (potential) cognitive strategies to solve them (Morand-Ferron et al., 2016; Pritchard et al., 2016). However, also less specific cognitive traits (e.g. associative learning) which are involved in behaviours in various contexts might be investigated, and detecting their role in specific fitness-determining behaviours such as predator avoidance, fighting, mating or foraging is important to understand correlations with fitness (Roth & Dicke, 2005; Morand-Ferron et al., 2016).

Studying variation in fitness-relevant traits is most meaningful in wild populations, i.e. in the environments to which organisms are adapted to and where selection is operating (Ellegren & Sheldon, 2008). Therefore, also individual variation in cognitive abilities that is linked with fitness outcomes should be studied in the wild or during short-term captivity which can be challenging (Thornton et al., 2014). Moreover, large sample sizes are needed to detect selection patterns, which imposes another challenge in the study of inter-individual differences of cognitive abilities and their link with fitness (Kingsolver et al., 2001; Dingemanse & Reale, 2005). Finally, studying several cognitive abilities and fitness traits, as well as non-cognitive factors simultaneously can help to detect patterns of covariation, fitness trade-offs and potential masking effects on the link between cognition and fitness, in order to fully understand the evolutionary consequences of cognition (Roth et al., 2010; Morand-Ferron et al., 2016).

Grey mouse lemurs as a study species

Studying cognitively sophisticated species such as corvids, cetaceans and primates is especially interesting in order to understand the evolution of cognition. Their close relatedness to humans make primates especially interesting for studying the fitness consequences of cognitive abilities. Many differences and similarities in cognitive abilities between humans and non-human primates have been studied in the past (e.g. Povinelli & Vonk, 2003; Herrmann et al., 2007; Matsuzawa, 2008). Primates have unusually large brains in relation to body size and most species live in complex social systems, which both have been linked to advanced cognitive abilities (Byrne & Whiten, 1988; Dunbar, 1998; Reader & Laland, 2002; Deaner et al., 2006; Dunbar & Shultz, 2007; Reader et al., 2011). However, life histories of primates are usually slow, which complicates the study of fitness outcomes. In addition, most primates are very difficult to habituate, thus testing them in cognitive tasks in the wild is challenging (but see Lührs et al., 2009; van de Waal et al., 2013; Huebner & Fichtel, 2015). Thus, until today nothing is known about potential fitness consequences of inter-individual variation in cognitive abilities of wild primates.

Grey mouse lemurs (*Microcebus murinus*) offer a unique opportunity to study cognition and fitness simultaneously in a wild primate population for several theoretical and practical reasons, however. These strepsirrhine primates are endemic to Madagascar and inhabit different forest habitats in the West and South of the island (Kappeler & Rasoloarison, 2003; Radespiel, 2006). They are small (60g), nocturnal primates that possess large brains relative to their body size (MacLean et al., 2009). Grey mouse lemurs forage solitarily, but individuals share sleeping nests and related females breed cooperatively in small sleeping groups (Eberle & Kappeler, 2006). The omnivorous ecological generalists have to respond flexibly to strong seasonal changes in food availability (Dammhahn & Kappeler, 2008b) while facing a high predation risk by various predators, including carnivores, snakes and owls (Goodman et al., 1993; Rasoloarison et al., 1995; Rahlfs & Fichtel, 2010; Fichtel, 2016). Thus, they face multiple ecological challenges under which species are likely to benefit from relevant cognitive abilities (Roth et al., 2010). Their cognitive abilities have been studied primarily in captivity (Joly et al., 2014; Kittler et al., 2018) but also in the wild with field experiments before the onset of this thesis (Lührs et al., 2009). As a practical advantage, mouse lemurs have one of the fastest life histories in primates as they reach sexual maturity in their first year of life and live on average only 2-3 years in the wild (Kraus et al., 2008; Hämäläinen et al., 2014), thus studying variation in fitness outcomes is feasible within few field seasons. Furthermore, wild grey mouse lemurs can be captured easily (“trap happiness”) (Kraus et al., 2008), which makes it possible to track individuals over the course of their lifetime. Also, individuals can be tested during several bouts of short-term captivity lasting only a few days, thereby minimizing the impact of captivity on behaviour and fitness consequences (Morand-Ferron et al., 2016). Moreover, grey mouse lemurs in the study population of Kirindy Forest have been captured and individually marked regularly since 1994 and individuals occur in large sample sizes (Eberle & Kappeler, 2002). Thus, grey mouse lemurs as study species and the wild population in Kirindy Forest offer many theoretical and practical advantages for a first study on fitness consequences of different cognitive abilities in a wild primate species.

Objectives and structure of this thesis

The aim of my thesis was to add to the small body of research investigating the adaptive value of cognitive traits for a better understanding of the evolution of cognition. Thus, with my thesis, I aimed to answer the questions: What causes and affects individual differences in cognitive performance? What is the structure underlying cognition, i.e., how are cognitive abilities related to one another? And ultimately: How are cognitive abilities linked with fitness outcomes? Therefore, I tested wild grey mouse lemurs in different, ecologically relevant cognitive abilities and linked individual variation in cognitive performance with different proxies of fitness. As highlighted before, I also tested the effect of potential non-cognitive factors affecting performance in the cognitive tasks and investigated the structure of individuals’ cognitive performance across tasks.

More specifically, in **Chapter 1**, I assessed inter-individual variation in cognitive abilities in five cognitive tasks and investigated the effect of individual characteristics and non-cognitive factors, like personality and motivation, on performance. Further, I investigated intra-individual variation in performance, i.e. the structure of cognitive performance and tested if there is a general factor explaining performance across different cognitive domains in grey mouse lemurs.

In **Chapter 2 and 3,** I linked individuals’ performance in the cognitive tasks with variation in fitness outcomes. In **Chapter 2**, I examined links between spatial learning ability and problem-solving efficiency and two fitness proxies: body condition change across the harsh dry season and survival. In **Chapter 3,** I focussed in detail on different fitness-related traits in grey mouse lemurs. I investigated variation in different fitness-related traits, i.e. body condition, hematocrit levels, and long-term cortisol concentration, that could be summarized into one factor. I tested if this physiological condition factor could be validated as a fitness proxy for a link with cognition. Moreover, I tested the link between cognitive performance in two problem-solving tasks, spatial learning and inhibitory control, and longevity, an important fitness proxy for grey mouse lemurs.

Finally, in the **General Discussion**, the findings of the three chapters are summarized and their implications for the ongoing study of the evolution of cognition are discussed. I draw overall conclusions and discuss limitations and future directions.

Chapter 1

Individual variation and the structure of cognitive performance in wild grey mouse lemurs (*Microcebus murinus*)

Franziska Hübner, Claudia Fichtel & Peter M. Kappeler

Author contributions:
FH, CF and PMK designed the study, FH conducted the experiments, analysed the data, and wrote the manuscript together with CF and PMK.

Abstract

Recent years have seen a surge of interest in inter-individual variation in cognitive abilities of a wide range of animal species. As a consequence, the underlying structure of cognitive performance, i.e., the question whether a general factor similar to the human general intelligence factor (*g*) can account for this variation in cognitive performance has received increased attention. However, evidence for *g* in animals has been mixed; perhaps because most studies were conducted in captivity and did not integrate individual characteristics and non-cognitive factors when testing for positive correlations among individuals' cognitive abilities in different tests and domains. We tested wild grey mouse lemurs (*Microcebus murinus*) in five cognitive tasks addressing different cognitive abilities and assessed effects of individual characteristics and non-cognitive factors on variation in cognitive performance. While grey mouse lemurs varied greatly in performance in the different tasks, we found no systematic effects of personality, body condition, motivation, age and sex on individual performance. Although task-directed motivation predicted performance in two problem-solving tasks, motivation measures were not correlated among tests, rendering a systematic effect on the covariation among cognitive performance measures unlikely. We found that performance in one cognitive task was generally a weak predictor of performance in any other task of our test battery, therefore providing no evidence for the existence of a general factor explaining cognitive performance in wild grey mouse lemurs. This first study of inter- and intra-individual variation in cognitive performance in a wild primate species therefore contributes several new insights for the comparative study of the evolution of general intelligence.

Keywords

Cognition, individual variation, motivation, personality, general intelligence factor, primate

Introduction

Recent research in animal cognition has experienced a shift from species differences to individual differences in cognitive abilities (Thornton & Lukas, 2012). Over decades, scientists have focused on comparative analyses of species differences in cognition and treated intra-specific variation as noise around a species mean (Thornton et al., 2014). However, to understand the evolution of cognitive abilities and potential fitness consequences thereof, studying individual differences in cognitive abilities is crucial. By acknowledging individual variation, several new, interesting questions can be asked: What causes and affects individual differences in cognition? What is the structure underlying cognition, i.e., how are cognitive abilities related to one another? And ultimately: How does cognitive performance relate to individual fitness outcomes?

Causes and confounding variables of individual variation in cognitive performance include factors like age, sex and personality, but also motivation, persistence, rearing condition and previous experience can potentially affect how animals perform in a given cognitive test (Thornton & Lukas, 2012; Rowe & Healy, 2014; Morand-Ferron et al., 2016; Dougherty & Guillette, 2018). Thus, controlling for these individual characteristics and non-cognitive factors while assessing individual differences in cognitive tasks is crucial to reliably compare cognitive performance across individuals. Especially the link between cognition and personality has received much attention recently (see meta-analysis by Dougherty & Guillette, 2018). Personality traits can co-vary with the way animals acquire, process and store information, i.e. they are linked to individuals' cognitive styles and result in speed-accuracy/flexibility trade-offs during learning (Carere & Locurto, 2011; Sih & Del Giudice, 2012; Griffin et al., 2015). Specifically, bold, fast-exploring, proactive animals are predicted to be faster but less accurate and flexible in learning a contingency, compared to shyer, slow-exploring and reactive individuals (e.g. Sih & Del Giudice, 2012; Mazza et al., 2018). However, the directions of these links between cognitive performance and personality traits were found to be highly variable across species (Dougherty & Guillette, 2018). Moreover, personality can also affect cognitive performance by increasing probabilities or rates of exposure with the task and the learning contingencies (Carere & Locurto, 2011; Morand-Ferron et al., 2016; Guillette et al., 2017). Especially in problem-solving tasks, where animals are tested with novel objects, and when testing wild, unhabituated individuals in cognitive tasks, investigating the effects of personality differences therefore appears crucial (Griffin & Guez, 2014; Morand-Ferron et al., 2016).

Besides identifying confounding factors of cognitive performance, quantifying whether individual variation in cognition is consistent across time and context is important (Griffin et al., 2015; Boogert et al., 2018). Low to moderate contextual and temporal repeatability could be demonstrated for different cognitive measures across species in a recent meta-analysis (Cauchoix et al., 2018) and appears to be widespread.

Individual differences in cognitive performance across different cognitive domains have been assessed using cognitive test batteries, which allow investigating how cognitive performances in different tasks relate to one another, shedding light on the structure of cognition. Subjects can either perform consistently across domains and tasks (domain-generality), or be good in one domain but fail in another (modularity) (Burkart et al., 2017). In human psychometric testing, performance across different cognitive domains is correlated and a single, general intelligence factor (g) explaining about 40% of variance in task performance can be extracted statistically (Plomin, 2001; Deary et al., 2010; Burkart et al., 2017).

Applying batteries of several cognitive tests to different species offers a possibility to illuminate the structure, evolution and function of cognition (Shaw & Schmelz, 2017). In animals, a g factor has been reported for some birds (Isden et al., 2013; Shaw et al., 2015; Ashton et al., 2018), rodents (Matzel et al., 2003; Galsworthy et al., 2005; Kolata et al., 2008; Light et al., 2010; Wass et al., 2012), dogs (Arden & Adams, 2016) and primates (Banerjee et al., 2009; Hopkins et al., 2014; Damerius et al., 2018). However, there are also studies that did not find evidence for the existence of a g in the same taxa or even the same species (birds: Boogert et al., 2011; Keagy et al., 2011; van Horik et al., 2018; rodents: Locurto et al., 2003, primates: Herrmann & Call, 2012; Herrmann et al., 2010).

For example, extensive studies of captive primates assessing various cognitive measures in large test batteries reported mixed results. Banerjee et al., (2009) reported evidence for g in cotton-top tamarins (*Saguinus oedipus*, N= 22) tested on 11 tasks on various cognitive abilities. In contrast, Herrmann et al., (2010) found no evidence for g in neither chimpanzees (*N*= 106) nor 2-year-old human children (*N*= 105) based on data obtained with the primate cognition test battery (PCTB) comprising 16 tasks from the physical and social domain (Herrmann et al., 2007). However, Hopkins et al. (2014) reported evidence for g in 99 chimpanzees based on their performance in 13 tasks in a modified version of the PCTB. Furthermore, Damerius et al. (2018) found evidence of g in orangutans (*Pongo pygmaeus wurmbii* and *Pongo abelii, N= 53*) based on five physical cognition tasks. This discrepancy in evidence for and against g indicates that drawing general conclusions across species and studies is currently difficult; also because studies differ in the nature and number of assessed tasks and domains, the degree to which confounding non-cognitive factors are controlled for, as well as in their general testing protocols and statistical methods (Burkart et al., 2017; Shaw & Schmelz, 2017; van Horik et al., 2018; Völter et al., 2018).

In addition, testing captive, highly enculturated subjects might result in a sampling bias, as individuals have a lifelong experience with humans, and rearing environments of captive animals might influence cognitive performance (Call & Tomasello, 1996; Würbel, 2001; van de Waal & Bshary, 2010; Thornton & Lukas, 2012; Sauce et al., 2018). Therefore, testing wild animals with different cognitive tasks is desirable; not the least because it also allows linking cognitive performance across tasks with fitness outcomes (Thornton et al., 2014). In a handful of studies on wild birds, applying test batteries, either directly in the wild or during short-term captivity, either revealed evidence of g (Isden et al., 2013; Shaw et al., 2015; Ashton et al., 2018) or not (Boogert et al., 2011; Keagy et al., 2011; reviewed in Shaw & Schmelz, 2017). Hence, additional studies measuring performance in several cognitive tasks in diverse wild animals are required to obtain a more profound understanding of the evolution of a general intelligence factor.

We therefore studied individual variation in cognitive performance in a wild primate species, the grey mouse lemur (*Microcebus murinus*). Mouse lemurs represent a suitable study species because they are small (60g), nocturnal, solitary, omnivorous Malagasy primates (Dammhahn & Kappeler, 2008b), and captive (Joly et al., 2014; Kittler et al., 2018) and wild individuals (Lührs et al., 2009; Henke-von der Malsburg & Fichtel, 2018; Huebner et al., 2018) have already been tested in different cognitive tasks and experiments. Wild individuals are "trap happy" (Kraus et al., 2008) and easily adapt to short-term captivity and cognitive testing (Huebner et al., 2018). Here, we explore cognitive performance in wild grey mouse lemurs using five cognitive tasks, while also investigating the potential effects of individual characteristics and non-cognitive factors, like feeding- and task-directed motivation and personality, on cognitive performance.

The five tasks were chosen to measure different ecologically relevant cognitive abilities, including the ability to learn a novel motor task, causal reasoning and spontaneous problem solving, inhibitory control, as well as spatial and reversal learning abilities (Table 1). We assessed associative motor learning during repeated and spontaneous innovative problem solving (Griffin, 2016; Griffin & Guez, 2014), which is of general ecological relevance as innovations in the wild allow animals to exploit new resources or to use existing resources more efficiently (Griffin & Guez, 2014; Reader & Laland, 2003). Inhibitory control, i.e. the ability to inhibit prepotent responses, is involved in various decision-making processes in asocial but also social contexts (Hauser, 1999; Amici et al., 2008; MacLean et al., 2014). Spatial learning and memory, i.e. the ability to remember the location of important resources, is crucial for mouse lemurs because they rely on sparsely distributed but predictable food resources during long and harsh dry seasons (Dammhahn & Kappeler, 2008a; Lührs et al., 2009). Finally, reversal learning paradigms test how quickly animals learn that a previously successful strategy is no longer rewarded, therefore assessing animals' behavioural flexibility, which plays an important role when environmental conditions change (Bond et al., 2007; Boogert et al., 2010). The aim of this study was, therefore, to quantify individual variation in cognitive performance in wild mouse lemurs while controlling for important non-cognitive factors and to test whether this variation can be explained by a single factor of general intelligence.

Methods

Study population and general procedure

This study was conducted at Kirindy Forest (CNFEREF), a dry deciduous forest in central Western Madagascar, at the research station of the German Primate Center. Grey mouse lemurs inhabiting a 10ha study area were regularly captured during the dry season from April to August in 2015 – 2017. Animals were captured with Sherman live traps, marked individually with subdermal microtransponders (Trovan Euro I.D., Frechen, Germany), sexed and aged (juveniles: less than 10 months old) based on morphometric data collected at the time of first capture (Dammhahn & Kappeler, 2008a).

In total, 97 individuals (44 females, 53 males, 63 juveniles, 34 adults) participated in the experiments of this study. For the experimental testing, animals were singly housed in the research station in 1m^3 cages containing natural branches and a sleeping box. Animals were kept for up to three consecutive nights before they were released again in the evening to their specific site of capture. Tests were conducted at night under dim red light. Subjects were rewarded with small pieces of banana in all cognitive tests and obtained a 1.5cm long piece of banana per night after the testing, while water was provided *ad libitum* (for more details on the housing and testing conditions see Huebner et al., 2018).

Subjects were first tested in two personality tests, followed by the cognitive test battery in a determined order (as presented below). As subjects could not be tested within one housing session (three nights), they were recaptured after a minimum of 10 days in their natural home range to continue with the experiments. Animals participated voluntarily in the cognitive tasks; if subjects refused to do so and/or did not appear and interact with the task apparatuses, tests were not counted and repeated on a subsequent night (Table 1 for total drop out numbers). All test sessions were video-recorded and analysed with the software BORIS (Friard & Gamba, 2016). We assessed inter-observer reliability with a second person naive to the research question scoring more than 10% of test sessions, which was excellent (intra-class correlation coefficient: food extraction task= 1, N= 10; string pulling task= 1, N=10; maze= 0.998, N= 10; Cohen's kappa: detour-reaching task= 0.87, N= 9).

Body condition as a proxy for food motivation

Since the body condition of small mammals reflects variation in energetic state (Schulte-Hostedde et al., 2005), we used it as a proxy for food-related motivation to participate in the food-rewarded tasks (reviewed in Griffin & Guez, 2014). We assessed subjects' body condition by calculating a body mass index (BMI) by dividing body mass (g) by bizygomatic breadth (mm), which reflects a reliable measure of linear body size in this species (Rasoloarison et al., 2000). Morphometric measures were taken as closely as possible to the date of testing, with the majority of measures being obtained within 4 weeks of testing.

Personality tests

Subjects' personality (measures of neophilia and activity) was assessed prior to habituation to the general test procedure, therefore representing individuals' baseline behaviour in an unfamiliar environment. An unknown open field arena (80x60x60cm) with four blind holes in the walls and two bigger entrances covered with mesh, was used for the personality tests (Dammhahn, 2012). A plotted grid with 12 cells helped to record the location of the subjects in the test arena. After a subject finished the personality test, the open field arena was cleaned with 70 % ethanol. The two personality measures were repeatable (see appendix).

Open field test

Open field tests offer a standardized tool to measure personality variation in animals (Réale et al., 2007; Dammhahn, 2012; Dall & Griffith, 2014). The test started with the release of a subject into the open field arena and lasted five minutes. During this time, we recorded individuals' activity as the total time (sec) subjects spent walking or climbing.

Novel object test

Animals' neophilia can be tested by measuring their response towards a novel object (Greenberg, 2003). The test lasted five minutes and started directly after the open field test with the introduction of a novel object (plastic toy ball, 2.5cm^3) into the open field arena. We recorded an animal's latency (sec) to contact this novel object as a measure of neophilia. If subjects did not contact the object within five minutes they were given the maximum latency of 300 seconds (i.e. capped latencies).

The cognitive test battery

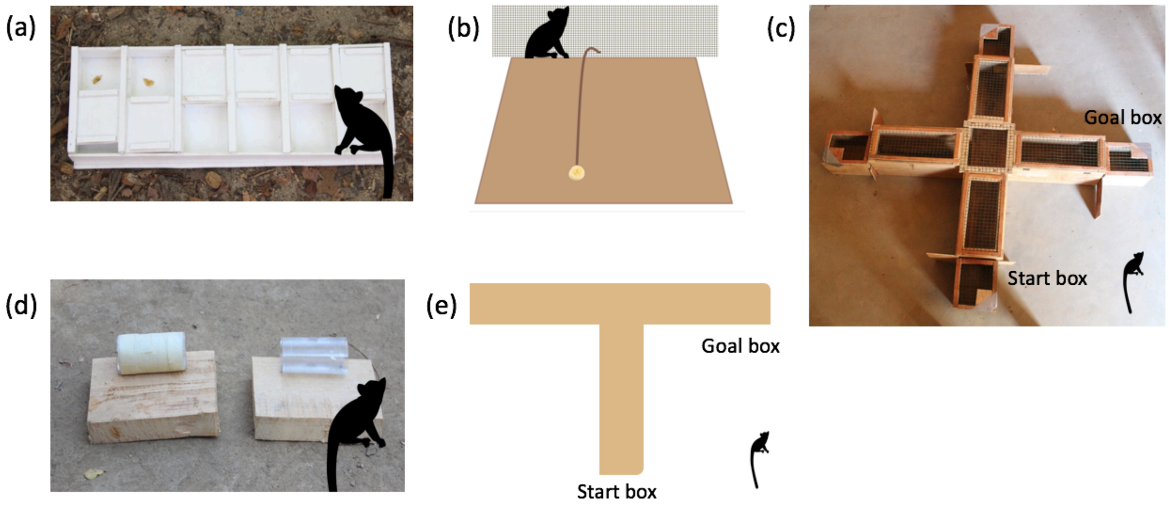

Figure 1: The cognitive test apparatuses.
(a) Food extraction task, (b) String-pulling task, (c) Plus maze, (d) Detour-reaching task, (e) T maze. Approximate proportional size of a mouse lemur is indicated by silhouette.

Food extraction task

During this novel motor task, subjects had to solve a novel problem repeatedly by removing sliding covers on six wells (5 x 4.5cm) of a small task box (Fig. 1), each containing a food reward. A small banana piece on top of the task box served as an initial incentive to interact with it.

During 20 minutes of testing, we recorded whether subjects had overall success in the task. For subjects that opened at least two lids, we recorded their solving time (s), i.e. the mean time they spent per successful opening after having opened the first lid. This measure of solving time addresses a subject’s efficiency in learning to apply the novel motor action as it correlates with subjects’ learning slopes in the food extraction task (Huebner et al., 2018). Moreover, individuals’ solving times were repeatable (Appendix) and therefore used as the main performance measure in this task.

To assess task-directed motivation during the experiment, we calculated individuals’ manipulation rates. We differentiated between manipulation rates before the first success and during repeated lid openings. Manipulation rates before the first success were calculated by dividing the time spent manipulating until the first successful opening by the latency until the first success, i.e. the time between the first contact of the box and first success. Manipulation rates during repeated solving were calculated by dividing the time spent manipulating the task apparatus after the first success by the time from the first success until the end of the experiment. This measure was then divided by subjects’ number of repeated door openings to control for differences in successes.

String-pulling task

In this task, a piece of banana attached to a string of 20cm length was positioned outside of the test cage, with the other end within reach of subjects through the cage wire (Fig. 1). During 20 minutes of testing, we recorded subjects’ latencies from the first attention to the reward until the successful pulling of the string. If subjects failed to pull the string and did not obtain the reward, we assigned maximal latencies. Solving latencies were repeatable (appendix). To assess task-directed motivation, we recorded subject’s attention towards the reward out of reach (the time when the head was oriented towards the reward/ string) and calculated attention rates by dividing the time spent with attention to the task by the solving latency.

Spatial learning

During the spatial learning task, subjects’ ability to remember the position of a food reward in a plus maze was tested. The maze consisted of four wooden arms (40x17cm; Fig. 1) with attached boxes (20x17cm) at each arm’s end. One of the boxes served as the starting point from where subjects were released into the maze, and either the arm to the left or the right served as goal box that contained a small banana reward (Huebner et al., 2018). To control for olfactory cues, large banana pieces were placed out of reach at the end of each maze arm, masking the smell of the actual reward. In order to avoid subjects from using own odour trails, the maze was cleaned with 70% ethanol after every third trial. Before testing, subjects had to pass a familiarization trial where they had to find rewards in all three arms of the maze.

Animals were then tested in 15 test trials. Each trial started with the release of the subject from the start box and ended with it obtaining the reward in the goal box. We recorded the number of errors made, i.e. the number of times animals entered an unrewarded maze arm, until reaching the learning criterion of finding the reward without errors in three consecutive trials or until the end of testing for animals that never met criterion, respectively. We graded the error scores to differentiate between the different levels of entering wrong arms. We assigned an error score of 1 if a subject entered the box at the end of an unrewarded arm, a score of 0.5 if a subject entered the arm but not the box at the end of the arm and a score of 0.25 when it entered a wrong arm with only part of the body.

Inhibitory control

We tested individuals' inhibitory control by assessing their ability to inhibit an ineffective prepotent response towards a food reward (MacLean et al., 2014). This detour-reaching task consisted of an open-ended transparent cylinder (20cm length, 6cm diameter, Fig. 1d, right cylinder) containing a clearly visible food reward in the centre. To control for odour cues that subjects might follow to retrieve the reward, small holes were made in the centre of the cylinder and the cylinder was cleaned on every fourth trial with 70% ethanol. For each trial, subjects were attracted with a small reward to one corner of their test platform before the cylinder was placed in the test cage so that subjects started to reach it from a central position at about 40 cm distance. In a familiarization phase, subjects had to retrieve a food reward out of an opaque cylinder (Fig. 1d, left cylinder) in five consecutive trials in order to start the test phase with the transparent cylinder. For the 10 trials of the test phase, we scored the number of erroneous trials subjects made when trying to first directly reach the reward before detouring to the open ends of the cylinder.

Reversal learning

In a T maze, subjects were tested in their ability to reverse a previously learnt reward location. The maze was similar to the plus maze but contained one arm less, i.e., subjects had to remember the location of a food reward either to the left or right of the starting arm (Fig. 1). The general testing procedure was similar to the plus maze used for spatial learning. Subjects habituated to the procedure during four familiarization trials. During the first familiarization trial, both arms were rewarded, followed by three trials where only one arm was rewarded and ended with the subject finding the reward in the correct location. Then the initial learning started with the rewarding scheme staying the same.

We tested animals in sessions of 10 trials and scored for each trial whether the subject was correct when it directly entered the rewarded arm. A subject reached the learning criterion when it directly entered the rewarded arm in nine of 10 consecutive trials (significantly exceeding chance level, binomial test, P= 0.022), either in a single test session or over two sessions. After a subject reached this learning criterion, it was subjected to the reversal learning test sessions the following night. To assure that subjects still remembered the rewarded location from the previous night, the reversal learning started with a repetition of this initial rewarding scheme. After subjects retrieved the reward during this repetition three times correctly, the rewarding scheme was reversed. Again, subjects were tested in sessions of 10 trials and had to reach the criterion of nine correct trials out of 10 consecutive trials. For each subject, we counted the number of trials needed to reach this criterion (minimum 10 trials) as a reversal learning score.

Table 1: Summary of cognitive tasks and respective performance measures, sample sizes and individuals' participation, presented in the order of testing.

Task	**Measure**	***N***	**addressing**	**Drop-outs ***
Food extraction	Solving time	77	Learning efficiency of novel motor action	1[a]
String pulling	Latency first success	97	Spontaneous problem solving, causal understanding	0
Plus maze	Errors until criterion	73	**Spatial learning**	13
Detour-reaching task	Errors in 10 trials	67	**Inhibitory control**	2
T maze	Trials until criterion	22	Behavioural flexibility during **reversal learning**	2

Sample sizes correspond to number of subjects for which personality was also measured.

*Individuals' participation: Animals that could not be tested because they did not participate voluntarily during task presentation were, if possible, retested during a subsequent night. We report final drop-out numbers here, i.e. animals that could not be retested or did not participate again during the next trial.

a) 97 animals participated in the food extraction task and for 77 individuals a solving time could be calculated. Only one animal could not be retested.

Statistical analyses

To determine whether individual characteristics and personality measures predicted performance in the cognitive tasks, we ran five different models with the respective cognitive performance measure as response variable. As age class and BMI were collinear, we first tested for a general age difference in cognitive performance in all tasks, using Mann-Whitney U tests. Since juveniles and adults did not differ in the main measures of cognitive performance, we combined the two age classes for further analyses. In all models, we implemented sex, BMI, neophilia and activity as predictors. We included an interaction between sex and neophilia to test for sex differences in the relationship between cognitive performance and personality (Dougherty & Guillette, 2018). However, these interactions were not significant and were therefore removed again in the final models. Moreover, to control for learning or solving opportunities in tasks where subjects interacted freely with the task during the time of testing (Griffin & Guez, 2016), we implemented the measures of task-directed motivation (manipulation and attention rate, respectively) as predictor variables in the models on performance in the food extraction and string-pulling task.

For the food extraction task, we first tested the effects of predictors on overall success in the task with a generalized linear model (GLM). Then we focused on solving time during repeated lid openings (our main performance measure in this task) which was log-transformed prior fitting as response in a general linear model (LM). To model the effects of predictors on solving latency in the string-pulling task and on the number of errors to reach criterion in the spatial learning task, we fitted Cox proportional hazards models (package survival in R: Therneau, 2015), treating maximal latencies for subjects that did not succeed (string-pulling task) and maximal errors for subjects that did not reach criterion (spatial learning) as censored observations. To model the effect of predictors on the proportion of erroneous responses in the inhibitory control task, we used a generalized linear model (GLM) with binomial error structure and logit link function. In this model, the number of failures and successes per individual was implemented with a two-column matrix as the response. To further investigate whether individual characteristics would predict task-directed motivation, we fitted two general linear models (LM) with attention rates in the string-pulling task or manipulation rates in the food extraction task as response variables and BMI, personality measures and sex as predictor variables.

Prior to fitting any model, we z-transformed covariates to facilitate interpretation of predictor estimates (Schielzeth, 2010). We checked the model assumptions ‘absence of collinearity’ using variance inflation factors (Fox & Monette, 1992; package car: Fox & Weisberg, 2011) and ‘absence of influential observations’ using dfbetas in all models (package survminer for Cox models: Kassambara & Kosinski, 2017). For the LM, we visually checked normally distributed and homogenous residuals and violation of proportional hazards for the Cox proportional hazards models. We always tested the full model against a null model containing the intercept only with an F-test (LM) or likelihood ratio test (GLM and Cox models).

As the sample size in the reversal learning task was low (N= 22), we used a Mann Whitney U test to test for sex differences in performance and Spearman rank correlations to test for correlations between individuals’ performance measures and BMI or personality scores.

To explore how performances in the different cognitive tasks were related to each other, we used Spearman rank correlations to test all pairwise correlations of the five tasks. To test whether a single general cognitive factor could explain performance across cognitive tasks, we performed a principal component analysis (PCA) with an unrotated factor solution and extracted principal components with an eigenvalue >1 (Burkart et al., 2017). We log-transformed the performance measures solving time in the food extraction task, and solving latency in the string-pulling task to achieve symmetrically distributed variables prior to PCA. We tested sampling adequacy of the correlation matrix used in the PCAs with the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett’s test of sphericity, which was considered appropriate with a KMO> 0.5 and Bartlett’s test of P< 0.05 (Budaev, 2010). As the sample size of 22 subjects in the reversal learning task was critically low for PCA (Osborne & Costello, 2004), we performed two PCAs, one containing the four cognitive tasks with high sample sizes, and one with all cognitive tasks and a lower sample size which we interpret with caution, as also KMO and Bartlett’s test criterions of sampling adequacy were not met. All analyses were conducted in R, v. 3.4.2 (R Core Team, 2017), only two-tailed tests were used, and the level of significance was set at 0.05.

Ethical note

All aspects of this study are in compliance with animal care regulations and applicable national laws of Germany and Madagascar. The Ministry for the Environment, Water and Forests of Madagascar, MINEEF and CNFEREF Morondava authorized research in Kirindy, and our research was approved by the relevant German Animal Use and Care committees and the animal welfare body of the German Primate Center (reference number E9-18).

Results

Effects of personality, motivation, body condition, age and sex on individual variation in cognitive performance

We found no significant differences between juvenile and adult subjects in any of the main cognitive performance measures (Mann-Whitney U test: solving time during food extraction: U= 725, P= 0.38, string-pulling latency: U= 1340, P= 0.26, spatial learning: U= 643, P= 0.33, inhibitory control: U= 478, P= 0.61, reversal learning: U= 28.5, P= 1).

Food extraction task

In total, 88% (N= 85) of subjects opened at least one lid in the food extraction task. We found a significant difference in task-directed motivation between successful and non-successful individuals in the food extraction task; solvers manipulated the apparatus significantly more often than non-solvers (Mann-Whitney U test: U= 29, P< 0.001, N= 97, Fig. A1). Furthermore, more juveniles than adults were overall successful in the task (Proportion test, X^2_1= 7.7, N= 97, P= 0.01). We did not find that sex, neophilia or activity predicted overall success in the task, but subjects with lower BMI were more likely to solve it (GLM, full null model comparison: X^2_4= 12.2, N= 97, P= 0.016, BMI: estimate± SE: -1.143± 0.421, P= 0.007, Table A1).

Solving times of subjects that opened lids repeatedly differed widely (mean± sd= 134± 161 s). We did not find an effect of sex, BMI, or the two personality measures on individuals' solving time, but manipulation rates predicted performance (LM, full null model comparison: $F_{5,71}$= 7.3, N= 77, P= <0.001, Table 2). Animals with higher manipulation rates had shorter solving times, i.e., they were quicker in opening the boxes repeatedly (Table 2).

Table 2: Results of the linear model: Effects of individual characteristics and non-cognitive factors on solving times in the food extraction task.

Predictor variable	**Estimate**	**SE**	**t**	**P**
Intercept	4.383	0.144	30.492	<0.001
Sex (male)	0.072	0.202	0.354	0.724
BMI[a]	0.037	0.1	0.371	0.712
Neophilia[a,b]	-0.062	0.096	-0.649	0.519
Activity[a]	0.037	0.095	0.39	0.697
Manipulation rate[a]	-0.547	0.092	-5.972	<0.001

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 77.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI± sd= 2.62± 0.4, neophilia= 176.2± 108.9 sec, activity= 200.8± 63.7 sec, manipulation rate= 0.06± 0.04.

[b] Neophilia was measured as latency to contact novel object in sec; low values correspond to neophilic behaviour

String-pulling task

Subjects varied in their latencies to solve the string-pulling task (mean± sd= 304 ± 384 sec) and 16 out of 97 individuals failed to solve the task. We found no effect of sex, BMI or personality, but attention rate predicted solving latencies (Cox proportional hazards model: full null model comparison: X^2_5= 32.9, N= 97, P< 0.001, Table 3).

Table 3: Results of the Cox proportional hazards model: Effects of individual characteristics on success latency in the string-pulling task

Predictor variable	**coeff**	**Exp(coeff)**	**SE(coeff)**	**z**	**P**
Sex (male)	0.228	1.256	0.26	0.876	0.381
BMI[a]	0.149	1.161	0.136	1.097	0.273
Neophilia[a,b]	-0.19	0.827	0.122	-1.558	0.119
Activity[a]	0.129	1.138	0.129	1.002	0.317
Attention rate to task[a]	0.655	1.925	0.13	5.038	<0.001

Positive coefficients indicate a higher hazard (here solving), i.e., shorter solving latencies. Exponentially transformed coefficients are the hazard ratios and give the effect size on the hazard of predictor variables. Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 97.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI± sd= 2.66± 0.43, neophilia= 181.64± 110.55 sec, activity= 199.6± 65.09 sec, attention rate= 0.58± 0.3

[b] Neophilia was measured as latency to contact novel object in sec; low values correspond to neophilic behaviour

Task-directed motivation

We found that subjects were consistent in their task-directed motivation within a task as manipulation rates until the first success and during the repeated lid openings correlated in the food extraction task (Spearman correlation: r= 0.372, N=77, P< 0.001). However, task-directed motivation was not consistent across tasks; subject's manipulation rates in the food extraction task and attention rates in the string-pulling task were not correlated (Spearman rank correlation, rho= 0.066, P= 0.572, N= 76, Fig.2). We neither found an effect of BMI, sex, neophilia and activity on subjects' manipulation rates in the food extraction task (full null model comparison: $F_{4,72}$= 0.72, P= 0.581, N=77, Fig.2, Table A2), nor on attention rates in the string-pulling task (full null model comparison: $F_{4,92}$= 2.23, P= 0.072, N=97, Table A3). Also, we failed to find an age difference in the two task-directed motivation measures (Mann Whitney U tests: manipulation rate: W=561, P= 0.76, attention rate: W= 1078, P= 0.457).

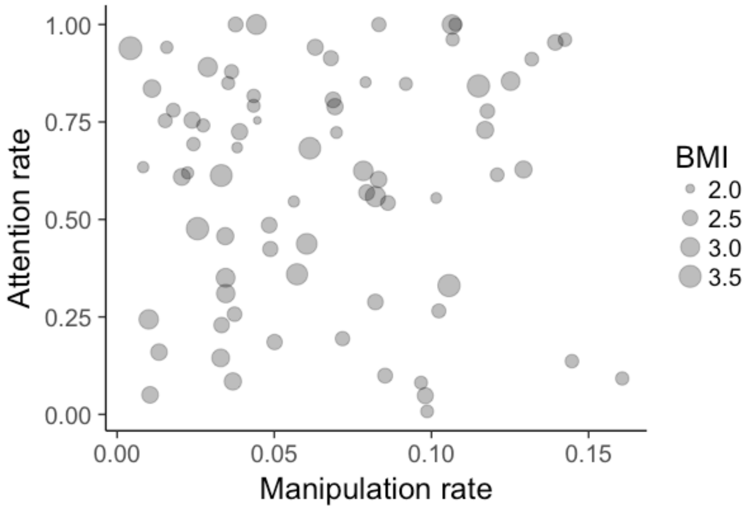

Figure 2: Relationship between the two measures of task-directed motivation: Attention rate in the string-pulling task and manipulation rates in the food extraction task. The size of circles reflects the BMI of subjects.

Spatial learning

Subjects differed in the number of errors they made until reaching the learning criterion in the plus maze (mean± sd= 14.2 ± 9.0). We found no link between individuals' sex, BMI, activity and neophilia and their learning performance in the plus maze (Cox proportional hazards model: full null model comparison: X^2_4= 1.96, N= 73, P= 0.744, Table 4).

Table 4: Results of the Cox proportional hazards model: Effects of individual characteristics on the number of errors until reaching the learning criterion in the plus maze

Predictor variable	**coeff**	**Exp(coeff)**	**SE(coeff)**	**z**	**P**
Sex (male)	-0.284	0.753	0.295	-0.961	0.337
BMI[a]	-0.089	0.915	0.141	-0.633	0.527
Neophilia[a,b]	0.064	1.066	0.154	0.418	0.676
Activity[a]	-0.142	0.868	0.164	-0.865	0.387

Positive coefficients indicate a higher hazard (here solving), i.e., shorter solving latencies. Exponentially transformed coefficients are the hazard ratios and give the effect size on the hazard of predictor variables. Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 97.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI± sd= 2.56± 0.41, neophilia= 173.41 ± 117.81 sec, activity= 207.57 ± 55.47 sec

[b] Neophilia was measured as latency to contact novel object in sec; low values correspond to neophilic behaviour

Inhibitory control

Subjects varied in the number of errors in the detour-reaching task (mean± sd= 3.2 ± 2.3). We found a sex difference in performance, with males making fewer errors, and a trend that subjects with a higher BMI made more errors (GLM: full null model comparison: X^2_4= 10.7, P= 0.03, N= 67, Table 5). We found no relationship between measures of individuals' personality and their performance in the inhibitory control task (Table 5).

Table 5: Results of the GLM: Effects of predictors on subjects' proportion of erroneous responses in the detour-reaching task.

Predictor variable	**Estimate**	**SE**	**z**	**P**
Intercept	-0.55	0.131	-4.184	<0.001
Sex (male)	-0.391	0.178	-2.193	0.028
BMI[a]	0.146	0.083	1.761	0.078
Neophilia[a,b]	-0.05	0.095	-0.526	0.599
Activity[a]	-0.073	0.091	-0.799	0.424

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 67.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI± sd= 2.63± 0.5, neophilia= 173.41 ± 117.81 sec, activity= 207.57± 55.47 sec

[b] Neophilia was measured as latency to contact novel object in sec; low values correspond to neophilic behaviour

Reversal learning

All mouse lemurs were able to reach the learning criterion in the reversal learning trials and differed in the number of trials needed to do so (mean± sd= 14.5 ± 4.1). We found no sex difference in the number of trials to meet criterion (Mann Whitney U test, U= 70.5, N= 22, P= 0.335). Also, performance did not correlate significantly with subjects' BMI and the personality measures activity and neophilia (BMI: r= -0.19, N= 22, P= 0.418; activity: r= -0.03, N= 21, P= 0.894; neophilia, r= 0.392, N= 21, P= 0.078).

Relationships between individual performances across tasks revealing the structure of cognition

Correlations of individuals' performances across tasks revealed that performance measures were not all positively correlated with each other, but that only solving time in the food extraction task and success latency in the string-pulling task correlated significantly positively with each other (Table 7). A principal component analysis with the test performance of the four main tasks (N= 52) revealed that not all cognitive performance measures loaded positively on the first principal component extracted (PC1 with Eigenvalue >1, KMO= 0.51, Bartlett's test: P= 0.005). PC1 contributed to 39.5% of the total variance in task performance (Table 8). Solving efficiency and latency in the string-pulling task loaded negatively on PC1, while spatial learning performance and detour-reaching performances loaded positively on PC1. When also adding reversal learning performance of subjects into the principal component analysis (N= 16, KMO= 0.41 Bartlett's test: P= 0.4), PC1 with an Eigenvalue >1 explained 36.3% of total variance and directions of variable loadings remained as before, with reversal learning performance positively loading on the first PC (Table 8).

Table 7: Spearman rank correlations of all cognitive task performances

	Food extraction task, Solving time	**String-pulling task**	**Spatial learning**	**Inhibitory control**
String-pulling task	**r= 0.27** **P= 0.016** **N= 78**			
Spatial learning	r= -0.11 P= 0.36 N= 63	r= -0.14 P= 0.24 N= 75		
Inhibitory control	r= -0.07 P= 0.63 N= 52	r= -0.22 P= 0.074 N= 67	r= -0.09 P= 0.49 N= 64	
Reversal learning	r= 0.02 P= 0.93 N= 17	r= -0.05 P= 0.84 N= 22	r= 0.36 P= 0.1 N= 22	r= 0.16 P= 0.5 N= 21

Significant correlation indicated in bold.

Table 8: Results of the principal component analysis

Task	Main tasks		Including reversal learning	
	PC1	PC2	PC1	PC2
Food extraction	-0.666	0.116	-0.429	0.312
String pulling	-0.657	-0.216	-0.592	-0.322
Spatial learning	0.308	-0.626	0.389	-0.670
Inhibitory control	0.171	0.740	0.524	0.505
Reversal learning	-	-	0.198	-0.308
Eigenvalue	1.256	1.080	1.346	1.105
% variance explained	39.5	29.2	36.3	24.4
N	52		16	

Discussion

We investigated the cognitive abilities of 52 wild mouse lemurs by assessing cognitive performance in four different main tasks targeting learning of a novel motor task, causal reasoning, inhibitory control and spatial learning. The aim of the first part of the study was to identify individual characteristics and non-cognitive factors that might have affected cognitive performance before investigating the structure of cognitive abilities in the second part.

The influence of individual characteristics on cognitive performance

Individuals differed greatly in performance across the cognitive tasks addressing different cognitive domains, which is an important prerequisite for investigating structures in cognitive performance (Shaw & Schmelz, 2017). However, we did not find a link between animals' personality, body condition or age class and their cognitive performance in any of the tasks. Also, personality traits did not predict how quickly animals learned during spatial and reversal learning (i.e. number of errors or trials until learning criterion). More neophilic and more active grey mouse lemurs were not quicker during spatial learning or less flexible during reversal learning. Our results do therefore not support the hypothesis of a speed-accuracy/flexibility trade-off that links personality types and cognitive styles during learning (Sih & Del Giudice, 2012; but see Dougherty & Guillette, 2018). Furthermore, neophilia and activity levels of subjects were not found to influence individuals' rates of exposure to task contingencies as task-directed motivation, which we measured directly in two of the tasks, was not predicted by these measures of personality. Thus, the assessed personality traits did not affect subjects’ performance in the tasks presented here.

However, task-directed motivation predicted performance in the two tasks in which subjects could also avoid engaging with the test apparatus (food extraction and string-pulling task). Subjects that were more interested in the tasks were also quicker in succeeding in the string-pulling task, and more manipulative individuals had a higher overall success and were more efficient in solving the food extraction task repeatedly. These findings support the notion that only animals that actively engage with a task are also likely to solve it. In motor tasks, in particular, manipulation of relevant and irrelevant parts of the task apparatus, often measured as persistence, correlated with learning of novel motor actions or other measures of problem-solving performance also in other species (great tits, *Parus major*: Cauchard et al., 2013; great tits and blue tits, *Cyanistes caeruleus*: Morand-Ferron et al., 2011; Indian mynas, *Sturnus tristis*: Griffin et al., 2014; common mynas, *Acridotheres tristis*: Lermite et al., 2017; pheasants, *Phasianus colchicus*: van Horik & Madden, 2016; spotted hyenas, *Crocuta crocuta*: Benson-Amram & Holekamp, 2012; grey squirrels, *Sciurus carolinensis*: Chow et al., 2016; redfronted lemurs, *Eulemur rufifrons*: Huebner & Fichtel, 2015; reviewed in Griffin & Guez, 2014).

In order to exclude the possibility that such task-directed motivation mediates correlations between performance measures (Shaw & Schmelz, 2017), it is important to account for the motivation to engage with a given task when analysing individual variation across cognitive tasks (as we did in the second part of the study). However, grey mouse lemurs were not consistent in their task-directed motivation across the two tasks, even though both tasks were food-motivated and conducted in the same night for the majority (87%) of subjects. Therefore, we think it is unlikely that general task-directed motivation mediated potential links between variation in individual cognitive performance across tasks. In other words, as individuals’ task-directed motivation in the food extraction and string-pulling task differed and was uncorrelated, it is unlikely that general motivation led subjects to perform similarly in the two tasks.

Neither personality, sex or age were found to affect subjects’ task-directed motivation, nor did body condition predict their motivation to engage with, or to pay attention to the task, perhaps suggesting that task-directed motivation itself might have a cognitive component. In principle, all subjects were generally motivated to eat the first freely accessible reward in the food extraction task, and thus seemed to be food-motivated, but not all subjects manipulated the boxes subsequently to extract more food. Because individual factors predicting task-directed motivation and persistence during task manipulations have not yet received much attention (but see Thornton & Samson, 2012; Lermite et al., 2017), it is presently difficult to place these results into a broader context.

Interestingly, inhibitory control performance differed between the sexes, with males exhibiting a better inhibitory control than females, which has not been reported for other species. Perhaps male grey mouse lemurs are better in inhibiting prepotent responses towards food because they experience more relevant situations in the wild. While female grey mouse lemurs hibernate during the food-scarce period, males continue foraging during this period when predation pressure also increases (Rasoloarison et al., 1995; Schmid, 1999; Rasoazanabary, 2006), making it perhaps more adaptive for males to be able to inhibit a response in potentially risky situations. Moreover, females have to accumulate body fat in the months before the dry season in order to be able to hibernate (Vuarin et al., 2013). As we conducted the experiments during this period of body mass accumulation, females might have been potentially keener to reach the food reward in the detour-reaching task.

Analyses of individual cognitive variation in wild subjects can be influenced by sampling biases as not all animals are willing to engage in tasks presented in the wild or to habituate to short-term captivity (Morand-Ferron et al., 2016; Shaw & Schmelz, 2017). In the current study, we did not have to limit our sample size to one sex in contrast to studies of wild birds (Boogert et al., 2011; Keagy et al., 2011; Shaw et al., 2015; but see Ashton et al., 2018), but we were able to test individuals of both sexes and different ages. Moreover, drop-out rates of subjects in the cognitive tasks were extremely low, meaning that almost all animals habituated to the test procedures and participated in the cognitive tasks (Table 1). In contrast to other studies in captivity focusing on highly trained lab animals, our subjects were naïve to general testing and had no prior experience with features of any of the tasks. Thus, we are confident that prior experience or sampling bias did not influence the performance of subjects across tasks.

To summarise, we could not identify a single non-cognitive factor that affected cognitive performance across tasks and that could potentially mediate correlations between task performances (Shaw & Schmelz, 2017). Therefore, in the second part of the study, we further investigated the underlying structure across cognitive performance in the different tests.

Is there a general factor explaining cognitive performance?

Only performance in the two problem-solving tasks was significantly positively correlated. For all other tasks, performance in one task was a weak predictor of performance in any other task of the test battery and subjects were not consistent in their performance across tasks. We applied a PCA in order to investigate whether there is a single factor explaining variation in cognitive performance, analogous to the general intelligence factor (g) in human psychometric testing. Performance in the food extraction and string-pulling task loaded strongly negatively on the first principal component, while performance in spatial and reversal learning and inhibitory control loaded positively on PC1. Thus, we could not find evidence of a general factor explaining variation in cognitive performance in these tests in grey mouse lemurs. Instead, individuals that scored high in the problem-solving tasks scored low in the other tasks addressing inhibitory control and spatial and reversal learning, suggesting the existence of a more modularized cognitive structure (Amici et al., 2012).

The weak but significant positive correlation between the cognitive measures of solving efficiency in the repeated food extraction task and spontaneous performance in the string-pulling task suggests that the two tasks address related cognitive abilities. In both tasks, subjects had to perform a novel motor action and to link this motor action with the food reward outcome. The use of problem-solving tasks or other operant motor tasks for the analysis of a potential g factor in animals has been criticized, as “problem solving” *per se* is a vague cognitive domain and performance in these tasks might be particularly prone to be influenced by task-directed or feeding motivation (Thornton et al., 2014; van Horik & Madden, 2016; Shaw & Schmelz, 2017) and related learning opportunities (Griffin & Guez, 2016). However, in our study, grey mouse lemurs’ problem-solving abilities were not influenced by personality traits, and performance in these tasks was unlikely to be predicted by a general factor of task-directed motivation (see above). Moreover, in the food extraction task, we tested subjects in their ability to apply a new motor action repeatedly. Since the measured solving time was correlated with individual learning slopes (Huebner et al., 2018), we think that performance in this task does indeed recflect cognitive abilities.

For performance in the string-pulling task, however, we cannot fully exclude the possibility that subjects solved the task by chance and not due to their ability to perform causal reasoning or associative learning (Thornton et al., 2014; Shaw & Schmelz, 2017; Jacobs, 2018) as we measured performance for the majority of subjects only once. However, performance in the task was repeatable for subjects that we retested (please see appendix), and the finding that performance in this task correlated with the performance in the motor learning task, while motivation to engage with the respective tasks did not, indicates that both tasks address similar cognitive abilities involved in understanding causal relationships, associative and motor learning. Moreover, also in the string-pulling task, subjects had to repeat the same motor action in order to obtain a reward because a single pulling action was only getting the reward closer, but not within immediate reach of the animal (see video or https://youtu.be/-90U3cFECdQ)

We found a weak positive, but nonsignificant correlation between individuals' spatial learning and reversal learning performance. As reversal learning was also tested in a spatial learning context in the less complex T maze, this trend demonstrates that not only behavioural flexibility, but also spatial learning abilities were important for remembering the reversed reward location. Reversal learning has been argued to not only address behavioural flexibility but also to reflect cognitive mechanisms involved in inhibitory control, as individuals have to first inhibit previously learned associations in order to learn the reversed contingency (Coppens et al., 2010; Izquierdo & Jentsch, 2012). In the mouse lemurs, performance in the detour-reaching task and reversal learning in the T maze tended to be weakly, albeit not significantly, positively correlated with each other despite the small sample size, which may indicate that similar cognitive mechanisms are involved. Positive correlations between detour-reaching and reversal learning performance have been also found in test batteries with wild birds (Shaw et al., 2015; Ashton et al., 2018).

When assessing the overall structure of cognitive performance, we found that individuals that performed well in the tasks that involved causal understanding and associative and motor learning during problem solving performed less well in inhibitory control and spatial and reversal learning. This contrasting performance within the same individuals might suggest a general trade-off in foraging strategies. Perhaps individuals that are less good at spatial learning, inhibitory control and behavioural flexibility during foraging compensate this handicap with better extractive and innovative foraging capacities. A similar suggestion has been made by van Horik et al. (2018), who found that pheasants that scored low in a detour-reaching task also had a high motor-related performance in two foraging tasks. More detailed studies on the link between foraging strategies and outcomes and cognitive abilities will be needed in the future to further explore this idea.

The few studies that also included problem-solving or motor task performance in their analysis of the general structure underlying variation in cognitive performance reported mixed results. Shaw et al. (2015) reported evidence of g based on testing of 16 North Island robins (*Petroica longipes*) in six tasks involving a motor task assessing motor learning skills. While there was evidence for g in spotted bowerbirds (*Ptilonorhynchus maculatus*) with a test battery including also problem-solving ability and motor skills (Isden et al., 2013), there was no evidence of g in a study of the closely related satin bowerbird (*Ptilonorhynchus violaceus*) (Keagy et al., 2011). The test battery of Keagy et al. (2011) also comprised problem-solving tests and tasks addressing behaviours closely related to males' natural display behaviour, but individual task loadings on the first component explaining 27.5% of variance were only partially positive. Similarly, there was no evidence for g in house sparrows (*Passer domesticus*) tested with a test battery including a motor task (Boogert et al., 2011). In the sparrows, results differed between test years, but in the 2010 sample, motor task performance and detour reaching performance loaded negatively on the first component that explained 37% of variance. There was also no evidence for a robust single factor comprising a broad variety of cognitive domains when testing pheasants in tasks addressing also novel motor skills (van Horik et al., 2018).

Importantly, only half of these studies on g in wild birds investigated the relationship between non-cognitive factors, like motivation and neophobia, and cognitive performance in these motor/ problem solving tasks before drawing conclusions on g (Keagy et al., 2011; Shaw et al., 2015; van Horik et al., 2018). Thus, these mixed findings further indicate that the different tasks and domains used, together with differences in the analyses, make it currently difficult to compare studies of g across species (Shaw & Schmelz, 2017). It therefore is currently an open issue whether problem-solving tasks should be included in studies of the structure of cognitive abilities. Conclusions about the presence and strength of g should be evaluated critically with regard to the specific measures used, the cognitive abilities they address, and whether non-cognitive factors are controlled for.

In conclusion, our study is the first to assess the structure of individual variation in cognitive performance in a wild primate species. The test battery we used allowed cognitive testing during short-term captivity of relatively large numbers of grey mouse lemurs. In the future, additional cognitive tests assessing additional cognitive domains and different tests for the same cognitive domain may help to characterize cognitive abilities in grey mouse lemurs more fully (Shaw & Schmelz, 2017; van Horik et al., 2018; Völter et al., 2018). We found no systematic effects of various individual characteristics and non-cognitive factors on performance in the cognitive tasks. In contrast to some other studies on captive primates, we did not find evidence for a general factor explaining variation in cognitive abilities. More carefully controlled and methodologically coordinated studies of various cognitive abilities in diverse species will be needed for a more systematic and comprehensive investigation of the evolution of general intelligence.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (awarded to P.M.K.: Ka 1082/33-1). We are very grateful to Bruno Tsiverimana, Léonard Razamanantsoa and all other members of the Kirindy research station team for their support in the field. Furthermore, we thank Beate Krauß and Lynne Werner for their help with the video scoring and the inter-observer reliability.

Appendix

Repeatability of personality measures activity and neophilia

Repeatability of personality measures from the open field and novel object test with grey mouse lemurs was reported to be low to moderate (Dammhahn, 2012; Henke-von der Malsburg & Fichtel, 2018) which might be explained by the testing of wild animals prior to human contact and habituation to general testing procedures. Therefore, we considered the personality measures as baseline behaviour prior to general testing and expected subjects to change their behaviour after associating general testing procedures with rewards. To demonstrate cross-contextual repeatability of neophilia and activity, we correlated them with behavioural measures during cognitive tasks, which addressed similar personality traits in comparable situations but were also conducted at the beginning of an individual's testing history. More specifically, our measure of activity in the open field test and the time animals needed to visit all three arms during the familiarization trial in the plus maze (see methods main text) correlated negatively (Spearman rank correlation, r= -0.266, N= 74, P= 0.022). Subjects that were more active in the open field test, were quicker in visiting all three arms in the plus maze.

The measure of neophilia in the novel object correlated with subjects' latency to contact the novel task apparatus in the food extraction task, a variable that is often directly used as personality measure (see Griffin & Guez 2014 for an overview on different variables used to assess neophobia or neophilia in studies on problem solving). Animals that were more neophilic in the novel object test, i.e. had a shorter latency to contact the novel object, were also faster to contact the food extraction box, resembling a novel object at the beginning of the food extraction task (intraclass correlation coefficient ICC= 0.263, N= 97, P= 0.004).

Repeatability of problem-solving performance

For the food extraction task and the string pulling task we were able to retest a subset of subjects after 10 to 30 days and calculated repeatability estimates using the rptR package (Stoffel et al., 2017). We fitted two linear mixed models for solving time in the food extraction task and solving latency in the string-pulling task (both variables log-transformed). To estimate 95% confidence intervals, we used parametric bootstrapping (1000 simulations) and likelihood ratio tests for significance testing. We calculated adjusted repeatabilities (Nakagawa & Schielzeth, 2010) by including individuals' number of tests, as some subjects were already tested a year before the repeatability testing, and test order of subjects as fixed factor in the models. Solving time in the food extraction task (r= 0.575, CI= 0.114– 0.916, N= 80 individuals, P= 0.043) and solving latencies in the string-pulling task (r= 0.535, CI= 0.175–0.867, N= 97 individuals, P= 0.005) were repeatable.

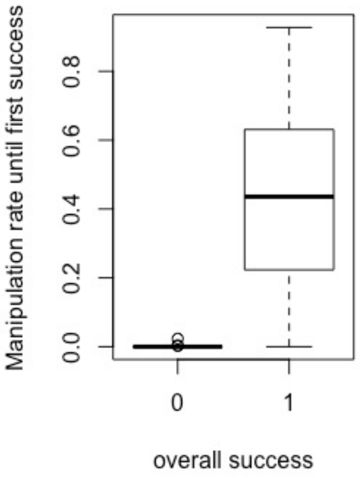

Figure A1. Differences in manipulation rates between subjects that succeeded in opening the first lid in the food extraction task and subjects that did not solve the task. Manipulation rates until the first success and total manipulation rates were used for subjects that solved (N= 85) and did not solve the task (N= 12), respectively.

Table A1: Results of the Generalized Linear Model (GLM) fitting the effect of individual characteristics on success probability in the food extraction task (success y/n)

Predictor variable	**Estimate**	**SE**	**z**	**P**
Intercept	2.862	0.715	4.003	<0.001
Sex (male)	-0.776	0.825	-0.941	0.346
BMI[a]	-1.143	0.421	-2.718	0.007
Neophilia[a,b]	-0.446	0.369	-1.21	0.226
Activity[a]	-0.543	0.373	-1.456	0.145

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 97.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI± sd= 2.67± 0.4, neophilia= 181.4± 110.7 sec, activity= 200± 65.4) sec

[b] Neophilia was measured as latency to contact novel object in sec; low values correspond to neophilic behaviour

Table A2: Linear model of the effects of individual characteristics on manipulation rates in the food extraction task

Predictor variable	**Estimate**	**SE**	**t**	**P**
Intercept	0.063	0.007	8.709	<0.001
Sex (male)	0.002	0.01	0.173	0.863
BMI[a]	-0.004	0.005	-0.732	0.466
Neophilia[a,b]	-0.005	0.005	-0.965	0.338
Activity[a]	0.003	0.005	0.653	0.516

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 77.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI± sd= 2.62± 0.4, neophilia= 176.2± 108.9 sec, activity= 200.8± 63.6 sec

[b] Neophilia was measured as latency to contact novel object in sec; low values correspond to neophilic behaviour

Table A3: Effects of individual characteristics on attention rates in the string-pulling task

Predictor variable	**Estimate**	**SE**	**t**	**P**
Intercept	0.671	0.046	14.444	<0.001
Sex (male)	-0.166	0.067	-2.495	0.014*
BMI[a]	-0.058	0.033	-1.747	0.084
Neophilia[a,b]	0.011	0.032	0.336	0.737
Activity[a]	0.034	0.032	1.076	0.285

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 97.

*Please note that the full null model comparison was not significant (see main text)

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI± sd= 2.66± 0.4, neophilia= 181.6± 110.6 sec, activity= 199.6± 65.1 sec

[b] Neophilia was measured as latency to contact novel object in sec; low values correspond to neophilic behaviour

Chapter 2

Linking cognition with fitness in a wild primate: Fitness correlates of problem-solving performance and spatial learning ability

Franziska Hübner, Claudia Fichtel & Peter M. Kappeler

Philosophical Transactions of the Royal Society B: 373: 20170295 (2018)

Special issue: "Causes and consequences of individual differences in cognitive abilities"

Author contributions:
FH, CF and PMK designed the study, FH conducted the experiments, analysed the data, and wrote the manuscript together with CF and PMK.

Abstract

Linking the cognitive performance of wild animals with fitness consequences is crucial for understanding evolutionary processes that shape individual variation in cognition. However, the few studies that have examined these links revealed differing relationships between various cognitive performance measures and fitness proxies. To contribute additional comparative data to this body of research, we linked individual performance during repeated problem solving and spatial learning ability in a maze with body condition and survival in wild grey mouse lemurs (*Microcebus murinus*). All four variables exhibited substantial inter-individual variation. Solving efficiency in the problem-solving task, but not spatial learning performance, predicted the magnitude of change in body condition after the harsh dry season, indicating that the ability to quickly apply a newly discovered motor technique might also facilitate exploitation of new, natural food resources. Survival was not linked with performance in both tasks, however, suggesting that mouse lemurs' survival might not depend on the cognitive performances addressed here. Our study is the first linking cognition with fitness proxies in a wild primate species, and our discussion highlights the importance and challenges of accounting for a species' life history and ecology in choosing meaningful cognitive and fitness variables for a study in the wild.

Keywords

Cognitive performance, fitness, survival, body condition, primate

Introduction

Observing animals around us, like a squirrel harvesting and caching nuts, it seems obvious that animals ought to benefit from cognitive abilities. Individuals of many species have to remember the location of food resources or shelters, respond flexibly to the presence of predators, potential mates or environmental changes, and could benefit from innovating new behavioural strategies in response to environmental change, for example. Cognitive abilities, i.e. the ability to acquire, process, store and respond appropriately to social and environmental information (Shettleworth, 2010), should therefore be associated with individual fitness benefits, so that individuals that learn faster, remember better, behave more flexibly, or innovate when confronted with new challenges, should on average also be in better body condition, produce more offspring and survive better. Nonetheless, not all animals have maximized cognitive capacities, but persistent individual differences in cognitive performance exist as higher cognitive performance is not only associated with fitness benefits but also with costs and therefore under selection (reviewed in Morand-Ferron et al., 2016, and see below). However, we still know little about the evolutionary forces and trade-offs that shape cognitive abilities as the links between them and fitness outcomes have been investigated in only a few species, and these studies revealed differing relationships (see below). Here, we contribute to this body of research by presenting results of the first study of the cognition-fitness links in a wild primate species.

Investigating fitness consequences of variation in cognitive abilities requires the study of both sets of variables in wild animals, which can be time-consuming and challenging for many practical reasons, especially for long-lived species (Cauchoix & Chaine, 2016; Morand-Ferron et al., 2016). In humans, intelligence has been linked to fitness-related traits like education, health and longevity (Plomin & Deary, 2015). However, evidence for the predicted positive relationship between cognition and fitness measures from animals is still rare, especially from the wild (Table S1). Among invertebrates, learning speed of bumble bee (*Bombus terrestris*) colonies correlated positively with colonies' overall foraging success (Raine & Chittka, 2008), but individual bumble bees' learning ability did not correlate with daily foraging performance, and bees with better learning abilities foraged for fewer days, indicating a (neuronal) cost of enhanced learning ability (Evans et al., 2017). In selected lab populations of fruit flies (*Drosophila melanogaster*), improved learning ability was also associated with a fitness cost and correlated with decreased larval competitive ability (Mery & Kawecki, 2003).

Among vertebrates, spatial learning accuracy in a maze correlated positively with reproductive success of captive rose bitterling males (*Rhodeus ocellatus*) in a sneaker role, but not in the dominant guarding role, the alternative male mating tactic in this fish species (Smith et al., 2015). Performance in problem-solving tasks, in which animals are presented with novel problems like artificial foraging tasks, was used as a measure of cognition in several studies of birds. However, this approach has recently been criticized because performance in problem-solving tasks is likely also affected by non-cognitive factors, and because the involved cognitive processes are not well defined (Cauchoix & Chaine, 2016; Rowe & Healy, 2014; Thornton et al., 2014). Nonetheless, in great tits (*Parus major*) (Cole et al., 2012; Cauchard et al., 2013; Preiszner et al., 2017) and house sparrows (*Passer domesticus*) (Wetzel, 2017), problem-solving performance correlated positively with measures of reproductive success, but problem-solvers also exhibited a higher probability of deserting their nests (Cole et al., 2012), suggesting associated fitness costs. Problem-solving performance of male satin bower birds (*Ptilonorhynchus violaceus*) in tasks closely related to natural display behaviour correlated positively with their mating success (Keagy et al., 2009, 2012). However, cognitive performance in a closely related species, the spotted bower birds (*Ptilonorhynchus maculatus*), did not correlate with male mating success when tested in a task battery addressing multiple cognitive abilities (Isden et al., 2013). Moreover, performance in cognitive tasks was not consistently related to song repertoire size, a predictor of various fitness-related traits, in song sparrows (*Melospiza melodia*): whereas reversal learning performance correlated positively with male song repertoire size, performance in two other cognitive tasks did not, and performance in a detour-reaching task was negatively related to song repertoire size (Boogert et al., 2011). Pheasant chicks (*Phasianus colchicus*) that were slow to reverse learned associations were more likely to survive for 60 days in the wild. Moreover, heavy pheasants that were quick in learning associations had improved survival, but for light animals, slow associative learners were more likely to survive (Madden et al., 2018). In Australian magpies (*Cracticus tibicen dorsalis*) group size was positively correlated with cognitive performance, and general cognitive performance in four different tasks predicted reproductive success in females (Ashton et al., 2018). Finally, wild male African striped mice (*Rhabdomys pumilio*) that were better in a long-term spatial memory task survived for longer, whereas female survival correlated negatively with the number of errors in a short-term spatial memory task (Maille & Schradin, 2016).

Thus, links between cognition and fitness outcomes have only been studied in a small number of wild vertebrate species, often focusing on members of one sex and on a single pair of variables. Furthermore, the differing results of these studies indicate that trade-offs of cognitive abilities and their links with fitness are likely to also depend on the study design such as the chosen cognitive measures, the conditions in which fitness measures are assessed, or individual characteristics like the sex or reproductive tactic of study subjects. Previous studies also demonstrated that, when studying the adaptive value of cognition, it is important to bear in mind that cognition is not a unitary trait, and that many different cognitive processes are involved in shaping a given behavioural outcome (Rowe & Healy, 2014). Moreover, cognition is involved in various different contexts, and what is beneficial in one situation can be disadvantageous in another (Rowe & Healy, 2014; Ten Cate, 2014). Furthermore, cognitive ability *per se* is likely to be associated with costs because neuronal tissue is energetically expensive and therefore also under selection (Kotrschal et al., 2013; Morand-Ferron et al., 2016). Hence, average individual cognitive performance in a particular test may not necessarily be closely and positively correlated with any fitness measure (Rowe & Healy, 2014), and detecting the underlying trade-offs is especially challenging in the wild (but see Cole et al., 2012). Nevertheless, stable inter-individual variation in cognitive abilities persists (Cauchoix et al., 2018), and relating it to variation in multiple fitness outcomes provides a reasonable starting point for a better understanding of the evolution of cognition (Thornton & Lukas, 2012; Thornton et al., 2014).

Primates stand out among mammals for their relatively large brains and social complexity, both of which have been linked to cognitive abilities (Byrne & Whiten, 1988; Dunbar, 1998; Reader & Laland, 2002; Deaner et al., 2006; Dunbar & Shultz, 2007; Reader et al., 2011). Because primates also have relatively slow life histories and wild populations do not readily cooperate in cognitive tests (but see Lührs et al., 2009; van de Waal et al., 2013; Huebner & Fichtel, 2015), nothing is known to date about potential fitness consequences of inter-individual variation in their cognitive abilities. Grey mouse lemurs (*Microcebus murinus*) are ideally suited among primates for such a study for several reasons, however. They are small (60g), nocturnal, solitary primates with large brains for their body size (MacLean et al., 2009). Grey mouse lemurs are omnivorous ecological generalists, responding flexibly to seasonal changes in food availability (Dammhahn & Kappeler, 2008b) while evading several types of predators (Rahlfs & Fichtel, 2010). In addition, juveniles have to complete growth and physiological preparations in time for several months of hibernation by the time they are about 6 months old (Schmid & Kappeler, 1998). Thus, grey mouse lemurs face multiple ecological challenges under which they are likely to benefit from relevant cognitive abilities (Roth et al., 2010). As a practical advantage, mouse lemurs can be easily captured with live traps, enabling us to bring them into a field laboratory for short-term cognitive testing before returning them to their natural home ranges. They also have one of the fastest life histories among primates, reaching sexual maturity within their first year of life and living on average for 2-3 years (Kraus et al., 2008; Hämäläinen et al., 2014), so that variation in fitness can be estimated within a few field seasons.

The specific aims of this study were, therefore, to test wild grey mouse lemurs in a problem-solving task and a maze and link test performance with fitness proxies. To this end, we measured problem-solving efficiency during repeated lid opening of an artificial foraging task and spatial learning by remembering a food location in a maze, and linked individual variation in test performance with body condition after the dry season, a strong predictor of survival and males' mating success (Eberle & Kappeler, 2004b), and with long-term survival. We expected performance in these two tasks to be ecologically meaningful and fitness proxies to be relevant because during the extended lean season that mouse lemurs face, spatial learning of available food resources and potential innovative foraging skills are likely to impact body condition and ultimately survival.

Methods

Study population and general procedure

This study was conducted at the research station of the German Primate Centre in the Forêt de Kirindy/ CNFEREF, a dry deciduous forest in central Western Madagascar (Kappeler & Fichtel, 2012). The study site is characterized by pronounced seasonality, with a 3-4 month hot wet season with high fruit and insect abundance followed by 8-9 months of a cool dry season with reduced food abundance during which mouse lemurs enter daily torpor or hibernation (Eberle & Kappeler, 2004b). Grey mouse lemurs living in a 10ha study area have been regularly captured and monitored since 1994 (Eberle & Kappeler, 2002). For this study, we used animals captured during monthly capture sessions between March and November, respectively, between 2015 and 2017. All animals were individually marked with subdermal micro transponders, sexed and aged (juveniles: <10 months old) based on morphometric data collected at the time of first capture (Dammhahn & Kappeler, 2008a).

For cognitive testing, individuals were kept at the research station in 1 m^3 cages containing a nest box and a testing platform. Tests were conducted at night and video-recorded under dim red light. Small pieces of banana served as reward in the tests. After testing, individuals were fed with a 1.5cm banana piece (minus the amount obtained in the tests) per night and water was provided *ad libitum*. After 1-3 nights in captivity, individuals were released in the evening at their specific site of capture and, if possible, recaptured after a minimum of 10 days for further cognitive testing. Cognitive tests were conducted at the beginning of the dry season, months before the start of the mating season, thus rendering it unlikely that individuals' fitness was affected by the few days in captivity. Testing subjects in captivity provided more controlled conditions and excluded potential threats from predators during the time of testing. Mouse lemurs were initially shy, but they habituated quickly and participated voluntarily in the experiments. We are therefore confident that testing under short-term captive conditions did not affect performance *per se*. Subjects were first tested with a food extraction task and then in a maze, either during three consecutive nights or after being recaptured. Because not all individuals could be recaptured with the same frequencies, sample sizes for the cognitive tests and fitness measures vary. Videos were analysed with the help of the software BORIS (Friard & Gamba, 2016). We assessed inter-observer reliability with a second person naïve to the research question scoring more than 10% of test sessions, which was excellent (intraclass correlation coefficient: food extraction task = 1, N= 10; maze = 0.998, N= 10).

Food extraction task

In the food extraction (FE) task, animals had to solve a novel problem by removing a sliding cover on each of the six wells (5 x 4.5cm) of a small box (6 x 12cm) in order to access a small piece of banana in each compartment (Fig. S1). Banana on top of the apparatus served as an initial incentive to start interacting with it. Subjects were presented with the task for a maximum of 20 min. If a subject did not appear on the test platform and interact with the box within 10 minutes (N= 16), the trial was not counted and repeated the following night. Fifteen of these subjects interacted with the box on the second attempt, resulting in a total sample size of 96 individuals for this task.

We recorded whether a subject opened at least one lid (general success: yes/ no), the total number of successes (0 to 6) and the latency from first contact with the box to first success. For subjects that interacted with the box but did not succeed, we recorded their total duration of testing, starting with the first contact with the box (i.e. capped latencies). Moreover, we measured an individual's solving time, i.e. the mean time a subject spent per successful opening after having opened the first lid, thus reflecting a subject's efficiency in repeatedly opening the lids of the novel motor task. For two subjects, we could not rate the total number of successes due to technical difficulties during testing. We were able to test part of the subjects repeatedly in the FE task with a time delay of 10 to 30 days and individuals' solving time was repeatable (intraclass correlation coefficient= 0.63, p= 0.044, N= 8; for other measures see Table S2).

Maze

In this spatial learning task, the ability of subjects to remember and retrieve the position of a food reward in a plus maze was tested. The maze consisted of four wooden arms (40cm x 17cm; Fig. S2) with attached boxes (20cm x 17cm) at each arm's end. One of the boxes served as the starting point from where subjects were released into the maze, and either the arm to the left or the right led to the reward (goal box). After successfully finding the reward in the goal box, the box was closed and rebaited before subjects were returned to the starting position and released again. To avoid the use of olfactory cues, big pieces of banana were placed out of reach at the end of every arm, thus masking the smell of the 2mm^3 reward inside the goal box. Each trial started with the release of the subject from the start box and ended with the subject consuming the reward in the goal box. After every third trial, the maze was cleaned with 70% ethanol in order to prevent individuals from using potential own odour trails as orientation cues. During an initial familiarization trial, all three boxes were rewarded, and subjects had to find all rewards to continue with the test trials. If subjects failed to find the food rewards within 10 min, testing was terminated and the familiarization trial was repeated on the following night. In total, 21 subjects did not complete the familiarization trial or stopped participating during the test session, but 12 of them could be re-tested on a subsequent day with eight subjects completing the test, resulting in a final sample size of 73 subjects.

During each of the 15 test trials, we recorded the number of errors subjects made, i.e., the number of times animals entered an unrewarded maze arm. More specifically, we rated a subject entering the box at the end of an unrewarded arm with a score of 1, entering a wrong arm with all four limbs, but not the box at the end, with a score of 0.5, and entering an arm with only part of the body with a score of 0.25. We defined a learning criterion, which was reached when a subject found the reward directly without any errors in three consecutive trials. For each subject, we determined whether it reached the learning criterion as well as the total number of errors it made until reaching the criterion or across all 15 trials.

Body mass index

To estimate body condition, which reflects variation in energetic state in small mammals (Schulte-Hostedde et al., 2005), we calculated a body mass index (BMI) by dividing body mass (g) by bizygomatic breadth (mm), the latter being a reliable measure of body size in this species (Rasoloarison et al., 2000). Because body mass fluctuates seasonally (Schmid & Kappeler, 1998), which may affect motivation to search for food rewards, we used individual's BMI measured up to two months prior testing and mean values for subjects that were measured several times in this time window. For a total of 44 subjects, we were also able to calculate the change in BMI between the end of the rainy season (mean of BMIs measured in March – May) and the end of the dry season (mean of BMIs measured in September – November) by subtracting the latter from the former.

Survival

We estimated individual survival by determining the number of days alive between birth and the date of last capture, truncating the study period in November 2017. Birth dates for all individuals were set at the modal birth date January 1 of the year of first capture for juveniles and one year earlier for subjects first captured as adults (see Eberle & Kappeler, 2004b). This second estimate is reliable because natal dispersals occurs within the first year of life (Schliehe-Diecks et al., 2012), and the probability of not capturing a natal individual within the first year of life is presumably extremely small. To define death operationally for individuals not recaptured for longer periods, we determined the 95th percentile of the frequency distribution of 10936 inter-capture intervals recorded between 1995 and 2017 as a cut-off point. Accordingly, study subjects were operationally considered dead if they were not recaptured within 161 days before 1 November 2017. In total, we could estimate survival for 84 individuals, excluding 11 juvenile males that presumably dispersed from the study area after their first test.

Statistical analyses

To evaluate the potential effects of individual characteristics, such as age, sex and body condition (which might proximately affect motivation), on performance in the cognitive tasks, we fitted multiple models with the respective measure of test performance as response and sex and BMI at the time of testing as predictor variables. We could not implement age class in these models, as BMI and age class were correlated and thus collinear. Therefore, to test for age differences in subjects' general ability to succeed in the FE task and to reach criterion in the maze, we ran proportion tests. To assess the effects of BMI and sex on subjects' probability to succeed in the FE task and on the probability to reach the learning criterion in the maze, we fitted Generalized Linear Models (GLM) with binomial error structure and logit link function. To model the effect on the number of successes and failures per individual in the FE task as response, we fitted a logistic Generalized Linear Mixed Model (GLMM) with individual identity included as random effect (R package *lme4*: Bates et al., 2015). We used a general linear model (LM) to fit the effect of sex and BMI on solving time (log-transformed) in the FE task. We used Cox proportional hazards models (package *survival* in R: Therneau, 2015) to model the effect on success latencies in the FE task and on the number of errors until criterion in the maze, treating maximal latencies for subjects that did not succeed (FE task) and maximal errors for subjects that did not reach criterion (Maze) as censored observations.

To determine whether an individual's performance in one task also predicted its performance in the other task, we used Spearman rank correlations for continuous measures of performance and Cohen's kappa coefficients for qualitative measures (success: yes/no in FE task, reached criterion: yes/no in Maze). We interpreted kappa values according to Landis and Koch (Landis & Koch, 1977).

To assess the effect of subjects' cognitive performance on their BMI change from the rainy to the dry season, we fitted LMs with BMI change as response. For the FE task, we implemented solving time (log transformed) as predictor and age and sex as control predictors. For the maze, we used the number of errors until criterion and the two control predictors. In both models, we first also tested the interactions between sex and performance measure and age class and performance measure. These interactions were not significant, but the full null model comparisons with the interactions and main effect removed were significant, and we therefore removed them from the model.

We used Cox proportional hazards models to fit the effect of cognitive performance, sex and age class on survival (in days). We implemented age class and sex as control predictors as these factors were previously shown to influence survival in mouse lemurs (Kraus et al., 2008). We fitted one model for the FE task with solving time (log transformed) and sex and age, and another model for the maze with the number of errors until criterion and the two control predictors. Again, interactions between test performance and age class and test performance and sex were removed from the models as they did not significantly explain individual survival. For all models, prior to fitting, we z-transformed covariates to a mean of zero and a standard deviation of 1 to facilitate interpretation of predictor estimates (Schielzeth, 2010). We checked the model assumptions "absence of collinearity" using Variance Inflation Factors (Fox & Monette, 1992; package *car* in R: Fox & Weisberg, 2011) and "absence of influential observations" using dfbetas in all models (package *survminer* in R for cox models: Kassambara & Kosinski, 2017). We controlled for the effect of potential outliers/ influential cases by comparing model results fitted with and without these observations but retained the complete dataset in all models. For LMs, we visually checked normally distributed and homogenous residuals and absence of overdispersion for the GLMM. For the Cox proportional hazards models, we checked the violation of proportional hazards. We always tested our full model against a null model containing the intercept only or just control predictors with an F test for general linear models and a likelihood ratio test for GLM, GLMM and Cox models. All analyses were conducted in R, version 3.4.2 (R Core Team, 2017). Level of significance was set at 0.05.

Results

Inter-individual variation in test performance

FE-task

Overall, 88% of 96 subjects successfully solved the FE task, i.e., they opened at least one of the six lids. Subjects varied in the total number of lids opened (mean± sd= 4.6± 2.15; CV= 46.74), their latency until the first success (mean± sd= 207± 325 sec; CV= 156.88) and solving time per successful opening after the first success (mean± sd= 134± 161 sec; CV= 120.53, Fig. 1a). An individual's BMI predicted its probability to open at least one lid (full null model comparison: X^2= 9.63, df= 2, P= 0.008; estimate± SE= -1.13± 0.41, z= -2.77, P= 0.006, N= 96,

Table S3), and also the total number of successes (full null model comparison: X^2= 6.73, df= 2, P= 0.035; estimate± SE= -2.26± 0.77, z= -2.94, P= 0.003, N= 94, Table S4): subjects with a lower BMI were more likely to solve the problem and opened more lids than subjects with a higher BMI. Moreover, subjects with a lower BMI had shorter latencies until first success (full null model comparison: X^2= 9.17, df= 2, P= 0.010, estimate± SE= -0.35± 0.12, z= -3.03, P= 0.002, N= 96, Table S5), but solving time per successful opening after the first success was not influenced by BMI (full null model comparison: $F_{2,73}$= 0.43, P= 0.655, estimate± SE= 0.11± 0.12, t= 0.92, P= 0.359, N= 76, Table S6). Significantly more juveniles than adults were successful (proportion test, X^2= 7.7, df= 1, P< 0.01, N= 97). Sex had no influence on any measure of performance in the FE task (probability of success: estimate± SE= -0.93± 0.82, z= -1.13, P= 0.258, N= 96, Table S3; number of successes: estimate± SE= -1.10± 1.38, z= -0.80, P= 0.423, N= 94, Table S4; latency first success: estimate± SE= -0.28± 0.23, z= -1.19, P= 0.234, N= 96, Table S5; solving time: estimate± SE= 0.10± 0.24, t= 0.43, P= 0.667, N= 76, Table S6).

Maze

In the maze, 71% of 73 subjects reached the learning criterion within the 15 test trials. Individuals varied in their number of errors until reaching the learning criterion (mean± sd= 14.24± 8.97; CV= 63.00, Fig. 1b), but juveniles and adults did not differ in their ability to reach the criterion (proportion test, X^2= 0.01, df= 1, P= 0.95, N= 73). BMI and sex did not influence performance and learning in the maze (probability of reaching criterion: full null model comparison: X^2= 1.68, df= 2, P= 0.431, BMI: estimate± SE= -0.17± 0.28, z= -0.59, P= 0.558, sex: estimate± SE= -0.73± 0.58, z= 1.26, P= 0.209, N= 73, Table S7; number of errors: full null model comparison: X^2= 0.81, df= 2, P= 0.667, BMI: estimate± SE= -0.09± 0.13, z= -0.71, P= 0.479, sex: estimate± SE= -0.20± 0.82, z= -0.69, P= 0.49, N= 73, Table S8).

Individuals' performance in the FE task and learning in the maze did not correlate between any performance measures (Table S9, Fig. S3). However, there was a tendency for successful animals in the food extraction task to be more likely to reach the learning criterion in the maze (Cohen's Kappa= 0.019, N= 71, Table S9).

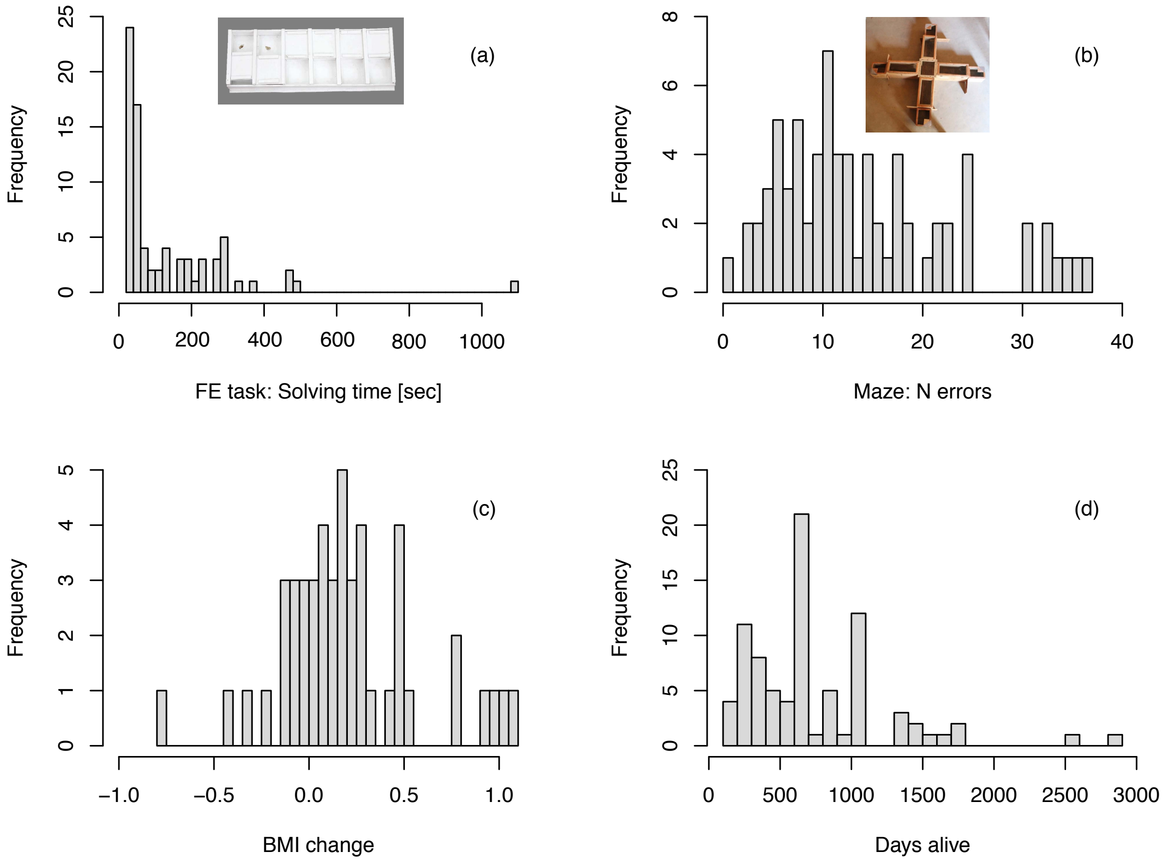

Figure 1. Inter-individual variation in performance in two cognitive tests and two fitness proxies. Depicted are histograms of the two main cognitive measures, (a) solving time of the FE task and (b) number of errors made until the learning criterion in the maze, and the two fitness proxies, (c) BMI change and (d) days alive.

Relationship between test performance and fitness proxies

Grey mouse lemurs varied in the two fitness proxies: BMI change during the dry season (mean± sd= 0.21± 0.37; CV= 176.19, Fig. 1c) and survival (mean± sd= 750.8± 499.1 days; CV= 66.48, Fig. 1d). Individuals' solving time, the measure of performance in the FE task that was not affected by body condition during the time of testing, predicted BMI change (full null model comparison: $F_{1,27}$= 4.742, P= 0.038). Animals that were slower in opening the lids after mastering it for the first time lost more body mass during the dry season (estimate± SE= 0.12± 0.05, t= 2.18, P= 0.038, Fig. 2a, Table S10). Moreover, BMI change was also affected by sex (females lost more body mass than males (estimate± SE= -0.48± 0.11, t= -4.35, P< 0.001, Table S9), but not by age (estimate± SE= -0.01± 0.11, t= -0.13, P= 0.900, N= 31, Table S10). Subjects' number of errors in the maze did not significantly predict their BMI change (full null model comparison: $F_{1,27}$= 3.88, P= 0.059), but there was a trend for animals that made more errors in the maze to experience a smaller change in BMI (estimate± SE= -0.12± 0.06, t= -1.97, P= 0.059, N= 31, Fig. 2b, Table S10).

Subjects' probability of survival was not predicted by their solving efficiency in the FE task (solving time: estimate± SE= 0.09± 0.15, z= 0.62, P= 0.534, N= 64, Fig. 2c, Table S11), whereas age class and sex predicted survival (full null model comparison: likelihood ratio test: X^2= 25.97, df= 3, P< 0.001). Specifically, juveniles and females had lower survival probabilities (age class: estimate± SE= 1.87± 0.44, z= 4.28, P< 0.001; sex: estimate± SE= -0.72± 0.31, z= -2.35, P= 0.019, Table S11). Mouse lemurs' survival probability was also not predicted by the number of errors they made in the spatial learning task (estimate± SE= -0.04± 0.16, z= -0.23, P= 0.824, N= 62, Fig. 2d, Table S12).

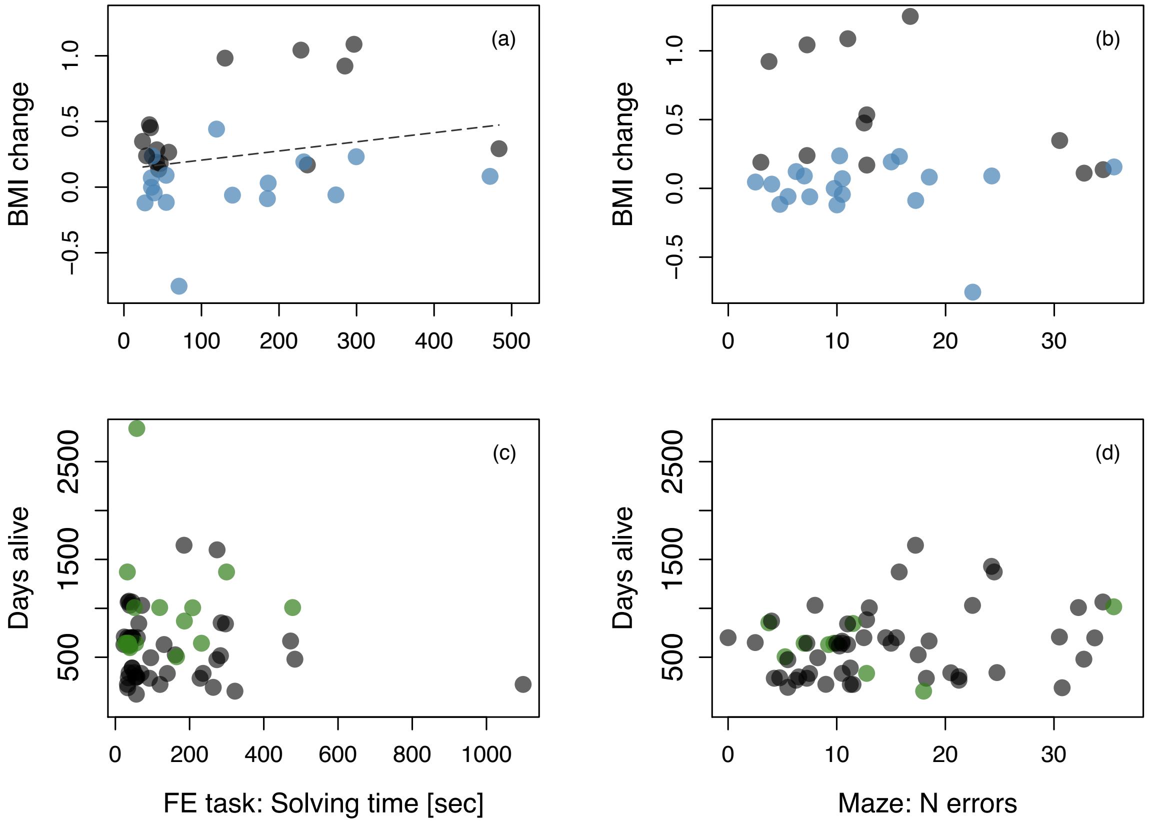

Figure 2. Relationship between BMI change and survival (number of days alive) and the cognitive performance measures, (a,c) solving time in the FE task and (b,d) number of errors in the maze. (a,b) BMI change: blue, males; grey, females; a positive BMI change corresponds to a decrease in BMI during the dry season, negative values reflect an increase in BMI from rainy to the end of dry season. (c,d) Survival: green, censored days alive for animals that are still alive; grey, dead animals.

Discussion

Individual variation in test performance

Our study contributes the first data on the cognition-fitness link for primates and established the feasibility of conducting cognitive tests with wild individuals during short-term captivity. We found that individual mouse lemurs varied in the chosen fitness proxies and in the measures of test performance in the two tasks, which is an important prerequisite for linking performance with fitness outcomes. Individuals' performance in the two tasks did not correlate, suggesting that there is no general factor underpinning performance in the two tasks, which presumably address different cognitive abilities (cf. Shaw & Schmelz, 2017). Importantly, solving time in the FE task and the number of errors in the maze were not affected by a subject's body condition, age or sex. Thus, we attempted to minimize the confounding effect of non-cognitive factors on individual test performance by linking variation in these performance measures with our fitness proxies in the second part of the study.

Inter-individual variation in test performance can be due to differences in cognitive abilities, but also to variation in motivation, personality, sex and age (Rowe & Healy, 2014; Morand-Ferron et al., 2016). Especially problem-solving tasks have been criticized as a measure of cognitive performance because differences in test performance might also be caused by variation in neophobia, persistence and prior experience or simply by chance (Thornton et al., 2014; van Horik & Madden, 2016; reviewed in Griffin & Guez, 2014), and because the specific cognitive processes underlying problem solving are not well defined (Healy, 2012; Rowe & Healy, 2014). We attempted to address this issue by testing the influence of several non-cognitive factors on test performance (see below), and by using a problem-solving design that allowed to test the repeated solving of the novel problem (Thornton et al., 2014). Thus, we not only measured performance during the criticized initial innovative problem solving, but also solving efficiency after the first successful opening of the artificial feeding box. Subjects with a low solving time efficiently and quickly opened the lids repeatedly after the first discovery of the novel solution and we suggest that they were able to do so because they quickly learned the new motor actions and associated them with the reward (cf. Griffin et al., 2014, Fig. S4).

Performance in problem-solving and other cognitive tasks can also be impacted by dimensions of individual personality, such as persistence, willingness to approach novel objects and speed to explore environments (Carere & Locurto, 2011; Dougherty & Guillette, 2018; reviewed in Morand-Ferron et al., 2016). To control for these potential effects, we assessed neophilia in a novel object task and general activity as well as exploration in an open field task. Details of these tests are beyond the scope of the present analysis and reported elsewhere (Huebner et al. in prep). The two cognitive performance measures (solving time in the FE task and number of errors in the maze) in the present study were not affected by these personality traits, however (Huebner et al. in prep). Thus, variation in these personality measures does not predict inter-individual variation in the measures of test performance in our study and are unlikely to mediate the correlation between solving efficiency in the FE task and body condition change as fitness proxy.

Moreover, when testing animals in food-rewarded tasks, controlling for motivation is equally important albeit difficult to operationalize. Differences in feeding motivation can be reduced in captivity by controlling access to food or water during a certain time window before testing animals, but this level of control cannot be achieved with wild animals. However, body condition may present a good proxy for the energetic state of wild individuals and, hence, their motivation to feed in the experiment. In line with the idea that “necessity drives innovation” suggesting that young, low-ranking individuals in poorer body condition are more likely to innovate (Clayton, 2004; Laland & Reader, 2003, but see Griffin & Guez, 2014; Reader & Laland, 2001), in the FE task, the initial and total number of successes, as well as first success latency were indeed affected by body condition at the time of testing, which differed widely between juveniles and adults. Juvenile mouse lemurs, which were tested during an important period of growth, had a lower BMI and appeared to be more motivated to solve the FE task than adults, which accumulated fat in the rainy season prior to testing. Yet, within a given age class, variation in BMI had no effect on test performance (unpublished data). In contrast to mammals, birds are limited in how much fat they can store (Witter & Cuthill, 1993), and motivation to feed (e.g. feeding latencies prior to testing, Sol et al., 2012), but not body condition (Griffin & Guez, 2014; Shaw, 2017) had an effect on problem-solving performance. Thus, lineage-specific constraints need to be considered and more comparative data are required for a more general assessment of the links between body condition and motivation in cognitive tasks.

Cognitive test performance and BMI change

Changes in BMI across the austral winter should reflect the ability of grey mouse lemurs to cope with the energetic challenges of a long cool dry season with reduced food availability (Dammhahn & Kappeler, 2008a; Schmid, 1999). Individuals exhibiting greater reduction in BMI lost disproportionately more fat reserves, indicating that they used more and/or acquired less energy than others between subsequent measures. Body condition at the end of the dry season is functionally relevant because it influences male mating success (Eberle & Kappeler, 2004b) and females' mating strategies (Huchard et al., 2012). Hence, this measure may also be meaningful for other small mammals or species experiencing strong environmental seasonality.

In our study, solving time in the FE task predicted BMI change, indicating a link between this specific measure of efficient, repeated problem solving and a fitness proxy. While necessity and motivation might drive initial innovations (Sol et al., 2012), after the initial discovery, associative learning and efficient reapplication of the new motor actions, as for example a novel behaviour to exploit new food resources, is crucial (Griffin & Guez, 2014). Especially under conditions where food resources are ephemeral, unpredictable and only seasonally available, innovation and efficient and swift associative learning of novel motor actions can be beneficial, as has been shown for several bird species (Sol et al., 2005; Sol et al., 2005). A previous field experiment with our study population revealed that mouse lemurs rapidly exploited new artificial feeding resources and swiftly learned changes in spatial arrangements (Lührs et al., 2009), suggesting that innovative foraging might be ecologically meaningful also under natural conditions.

Performance in the maze was not linked to BMI change in this study. We chose this test because we expected a positive correlation between an animal's ability to remember a food location in the maze and its ability to remember and find natural food resources, which, in turn, should affect body mass dynamics. Failure to demonstrate this link could be due to two reasons. First, females hibernate for several months during the dry season, whereas males only enter short daily torpor bouts (Rasoazanabary, 2006; Schmid, 1999). Thus, remembering food locations may not be subject to strong selection in females. In contrast, males feed on tree gum and sugary secretions of colonial invertebrates during the lean dry season, which are both patchily distributed, so that remembering the location of these food resources might be beneficial. Despite this sex difference in natural foraging ecology, the effect of test performance on BMI change did not differ between males and females, however. Second, variation in motivation and explorative behaviour might have influenced the number of errors in the maze. There was indeed a trend indicating that subjects making more errors in the maze experienced smaller BMI changes. However, this trend could not be explained by the current BMI, our proxy for feeding motivation. Also, subjects were highly motivated to participate in all food-rewarded tasks, and we never observed any animal rejecting offered food. Unfortunately, we could not control for individual variation in exploratory behaviour during the trials in the maze. Imposing a cost for exploring the environment, as for example in the Morris water maze (D'Hooge & De Deyn, 2001), might allow to evaluate this possibility in a future study.

Cognitive test performance and survival

Grey mouse lemur's solving time in the FE task or spatial learning performance in the maze did not predict their subsequent survival in the current study. One possible explanation for our failure to find a relationship between these performance measures and survival might involve a lack of statistical power, even though our sample sizes were larger than those in most previous studies of primate cognition. However, we found a significant correlation between BMI change and performance in the FE task, and several recent studies with even smaller sample sizes could demonstrate a link between cognitive measures and fitness proxies (Table S1). Thus, it is possible that mouse lemurs' survival might not be predicted by the specific cognitive abilities addressed here.

While the two tests measure cognitive performances that ought to impact survival via body condition, mouse lemur survival is probably impacted more profoundly by predation risk. Among primates, mouse lemurs are exposed to one of the highest predation rates (Fichtel, 2012) and are preyed upon by various carnivores, owls, snakes and even another lemur species (reviewed in Fichtel, 2016). Predator avoidance has been shown to be linked with survival in striped mice: female survival was predicted by a faster response to predator stimuli, and male survival co-varied positively with better long-term spatial memory of shelter locations. In contrast, female striped mice that made more errors in a maze testing short-term memory survived longer, and overall survival was not linked to performance in the spatial memory task (Maille & Schradin, 2016), indicating that even when linking predator avoidance performance with survival, the direction of these links are not necessarily as predicted.

Because grey mouse lemurs are nocturnal, certain anti-predator tactics, such as vigilance and subsequent fleeing to a distant shelter, are not effective (Fichtel, 2016). Instead, grey mouse lemurs tend to freeze after detecting a predator (Rahlfs & Fichtel, 2010), a behaviour that is more difficult to address in a laboratory cognitive task. Thus, a species' sensory ecology and their actual specific behaviours in fitness-relevant contexts needs to be taken into account when choosing appropriate cognitive tests and fitness proxies (Cauchoix & Chaine, 2016).

Conclusions

Our study indicates that links between experimental measures of cognitive test performance and fitness proxies of wild animals are not necessarily direct and easy to assess and interpret. It is essential to appreciate a species' life history and ecology in studying how selection shapes certain cognitive abilities, not only with regard to study design, but also with respect to the complex interactions among cognitive performance and confounding factors like personality, motivation, age and sex differences. Similarly, fitness proxies have been notoriously difficult to measure in behavioural ecology, especially when egg-counting is not an option, and this and most other mammal species offer examples for the practical challenges of identifying and operationalizing meaningful fitness proxies. Thus, more comprehensive study designs than bivariate correlations will be required in the long term to broaden our understanding of the evolutionary mechanisms underlying species-specific adaptations in cognitive abilities and their intra-specific variation.

Acknowledgments

We are very thankful to Bruno Tsiverimana, Léonard Razamanantsoa and all other members of the Kirindy research station team for their support in the field. Moreover, we thank Henning Lahmann for the administration of the long-term data, Lynne Werner for scoring videos for the inter-observer reliability and Roger Mundry for statistical advice. Finally, we thank Alex Thornton and two anonymous reviewers for their constructive comments.

Supplemental material

Figures

Figure S1: The food extraction task: Body width of a mouse lemur corresponds to the width of one compartment (5 x 4.5cm).

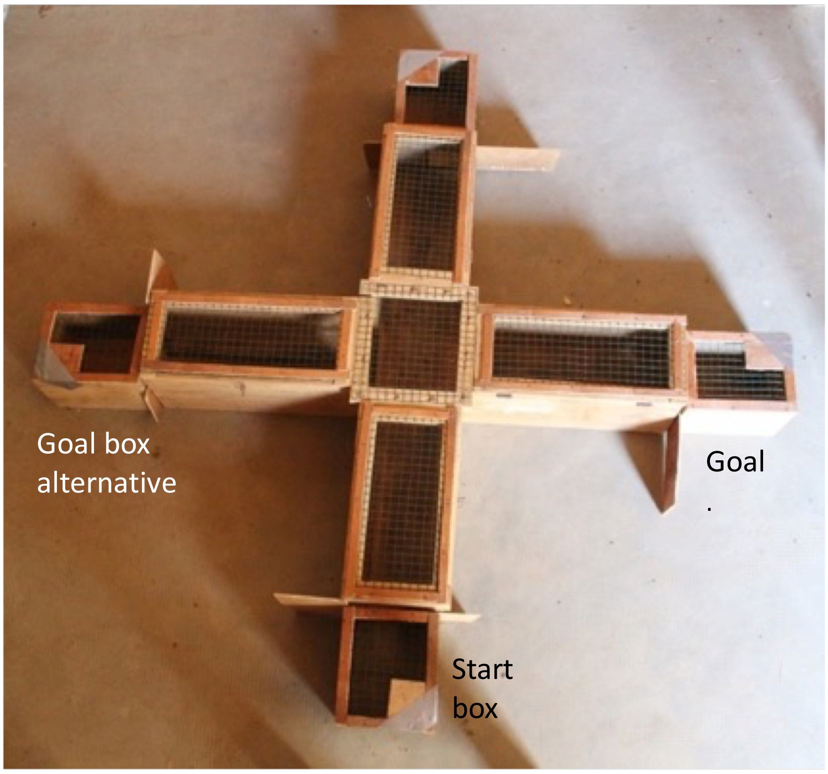

Figure S2: The maze: Body size of a mouse lemur corresponds approximately to one quarter of the start box (20cm x 17cm).

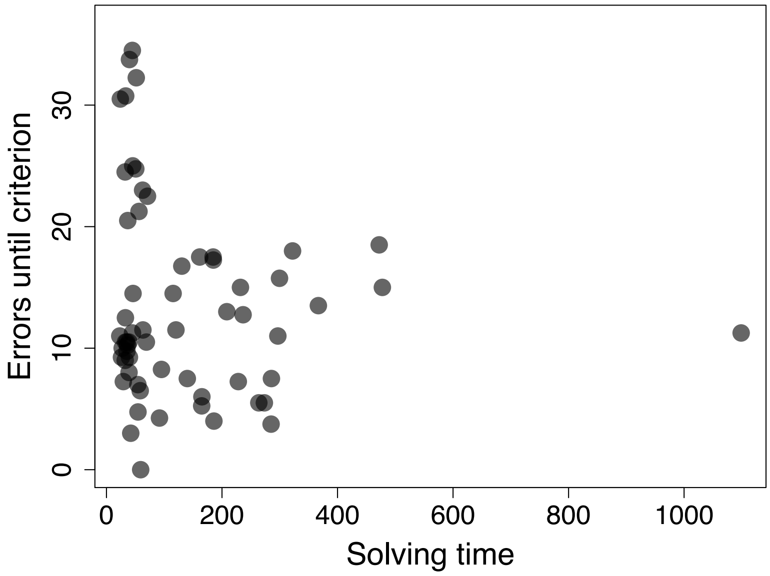

Figure S3: Correlation between the two main measures of cognitive performance: solving time (in seconds) in the FE task and number of errors until criterion in the maze.

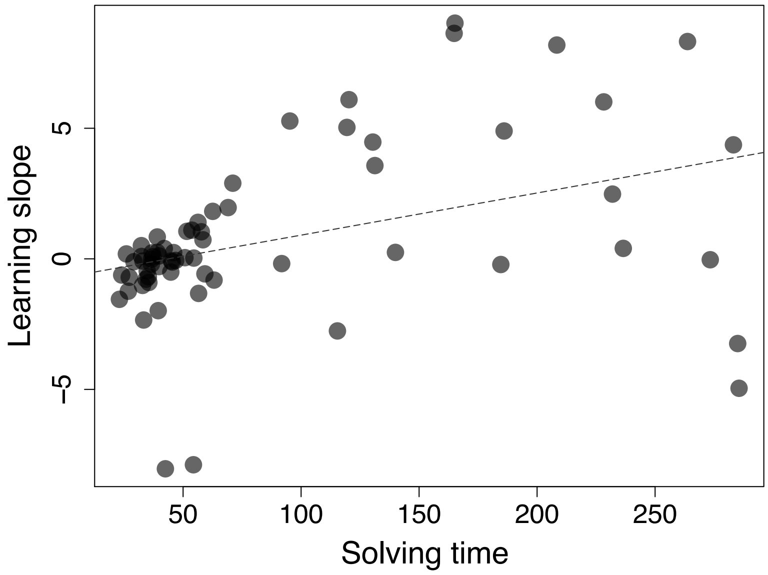

Figure S4: Correlation between individual learning slopes and mean solving time (in seconds) for subjects that opened at least five lids in the FE task. Learning slopes were calculated from individual regression lines of successive latencies until lid openings from first success until fifth or sixth success (i.e. time intervals between successes). Negative slopes reflect a decrease in solving latencies and suggest learning across lid openings. Spearman rank correlation (r= 0.46, S= 25784, P< 0.001, N= 66) revealed that individuals' mean solving time and learning slopes correlated positively, thus supporting the notion that for subjects with low solving times, learning is involved during the repeated opening of lids in the FE task and that individuals' mean solving times are an adequate measure to compare among subjects that differed in the number of lids opened.

Tables

Table S1: Overview of studies linking cognitive performance and fitness proxies

Species	**Cognitive performance**	**Fitness proxy**	**Sample size**	**Relationship**	**Fitness measured in**	**Reference**
Bumble bee, *Bombus terrestris*	Associative learning	Foraging success	12 colonies	positive	wild	Raine & Chittka 2008
Bumble bee, *Bombus terrestris*	Associative learning	Lifetime foraging performance	85	negative	wild	Evans et al. 2017
Rose bitterling, *Rhodeus ocellatus*	Spatial learning	Reproductive success	16 males	positive; depending on mating tactic	captivity	Smith et al. 2015
Great tit, *Parus major*	Problem solving	Clutch size	368 females	positive	wild	Cole et al. 2012
Great tit, *Parus major*	Problem solving	Nest success	368 females	negative	wild	Cole et al. 2012
Great tit, *Parus major*	Problem solving	Adult survival	698	none	wild	Cole et al. 2012
Great tit, *Parus major*	Problem solving	Fledgling number, clutch size, nestling survival	26 pairs	positive	wild	Cauchard et al. 2013
Great tit, *Parus major*	Problem solving; 2 tasks	Hatching success, Fledgling number	55 pairs	positive for 1 problem-solving task	wild	Preiszner et al. 2016
Great tit, *Parus major*	Problem solving; 2 tasks	Clutch size	55 pairs	none	wild	Preiszner et al. 2016
House sparrow, *Passer domesticus*	Problem solving	Nestling survival	80	none for females, positive for males (N= 41)	wild	Wetzel et al. 2017
Satin bower bird, *Ptilonorhynchus violaceus*	Problem solving, 2 tasks	Mating success	33 (25) males	positive	wild	Keagy et al. 2009
Spotted bower bird, *Ptilonorhynchus maculatus*	PC score* from 6 tasks (Motor task,	Mating success	11	none	wild	Isden et al. 2013

	color and shape discrimination, reversal learning, spatial memory)					
Australian magpie*, Cracticus tibicen dorsalis*	PC score* from 4 tasks (inhibitory control, associative learning, reversal learning, spatial memory)	Number of clutches and fledglings per year	22 females	positive	wild	Ashton et al. 2018
African striped mouse, *Rhabdomys pumilio*	Spatial memory	Survival until breeding season	20 males, 22 females	positive for males, negative for females	wild	Maille & Schradin 2016

* Scores from principal component analysis used for the further analysis

Table S2: Results of the repeatability tests for measures of the FE tasks

Performance measure	**Test**	**Result**	**Sample size**	**Interpretation**
Success yes/ no	Cohen's kappa	Kappa= 0.42	13	Moderate agreement
Latency success	Intraclass correlation	ICC= 0.34	12	Poor agreement
Solving time	Intraclass correlation	ICC= 0.63	8	Good agreement

Subjects were tested in the same task with a delay of 10 to 30 days. On the group level, subjects improved in performance: Latency to success decreased by 205± 500 sec (mean± sd), solving time decreased by 72± 65 sec (mean± sd). Interpretation of Cohen's kappa and intraclass correlation coefficients according to Hallgren 2012.

Table S3: Food extraction task: results of the Generalized Linear Model (GLM) fitting the influence of BMI on success probability (success y/n)

Predictor variable	**Estimate**	**SE**	**z**	**P**
Intercept	2.83	0.70	4.03	<0.001
BMI[a]	-1.13	0.41	-2.77	0.006
Sex (male)	-0.93	0.82	-1.13	0.258

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 96.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI (sd)= 2.66 (0.39).

Table S4: Food extraction task: results of the Generalized Linear Mixed Model (GLMM) testing the influence of BMI on individuals' number of successes

Predictor variable	**Estimate**	**SE**	**z**	**P**
(Intercept)	5.70	1.45	3.92	<0.001
BMI[a]	-2.26	0.77	-2.94	0.003
Sex (male)	-1.10	1.38	-0.80	0.423

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 94.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI (sd)= 2.66 (0.39).

Table S5: Results of the Cox proportional hazards model fitting the effects of body mass index on latency to solve in the food extraction task

Predictor variable	**coeff**	**Exp(coeff)**	**SE(coeff)**	**z**	**P**
BMI[a]	-0.35	0.71	0.12	-3.03	0.002
Sex (male)	-0.28	0.76	0.23	-1.19	0.234

Positive coefficients indicate a higher hazard (here solving), i.e., shorter solving latencies. Exponentially transformed coefficients are the hazard ratios and give the effect size on the hazard of predictor variables. Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 96.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI (sd)= 2.66 (0.39).

Table S6: Food extraction task: results of the linear model testing the effect of body mass index at time of testing on individuals' solving time

Predictor variable	**Estimate**	**SE**	**t**	**P**
Intercept	4.35	0.17	25.14	<0.001
BMI[a]	0.11	0.12	0.92	0.359
Sex (male)	0.10	0.24	0.43	0.667

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 76.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI (sd)= 2.61 (0.37).

Table S7: Maze: results of the Generalized Linear Model (GLM) fitting the effect of predictors on subjects' probability to reach the learning criterion

Predictor variable	**Estimate**	**SE**	**z**	**P**
Intercept	1.34	0.45	2.98	0.003
BMI[a]	-0.17	0.28	-0.59	0.558
Sex (male)	-0.73	0.58	-1.26	0.209

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 73.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI (sd)= 2.56 (0.35).

Table S8: Maze: results of the Cox proportional hazards model fitting the effect of predictors on individuals' number of errors until reaching the learning criterion

Predictor variable	**coeff**	**Exp(coeff)**	**SE(coeff)**	**z**	**P**
BMI[a]	-0.09	0.91	0.13	-0.71	0.479
Sex (male)	-0.20	0.82	0.28	-0.69	0.488

Positive coefficients indicate a higher hazard (here reaching the learning criterion), i.e., fewer errors. Exponentially transformed coefficients are the hazard ratios and give the effect size of predictor variables on the hazard. Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 73.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean of BMI (sd)= 2.56 (0.35).

Table S9: Relationships between performances in the maze and in the food extraction tasks tested with Spearman rank correlations and Cohen's Kappa tests

	FE: Latency success	**FE: N of successes**	**FE: solving time**	**FE: success y/n**
Maze: Errors until criterion	R_S= 0.13 P= 0.27 N= 71	R_S= -0.08 P= 0.49 N= 69	R_S= -0.08 P= 0.56 N= 61	/
Maze: criterion y/n	/	/	/	Cohens Kappa= 0.019, N= 71

Table S10: Results of the linear models (LM) fitting the effects of test performance in food extraction task and maze on BMI change from the rainy to the end of dry season

	Predictor variable	**Estimate**	**SE**	**t**	**P**
Model 1: Food extraction N= 31	Intercept	0.48	0.08	5.73	<0.001
	Solving time[a]	0.12	0.05	2.18	0.038
	Sex (male)	-0.48	0.11	-4.35	<0.001
	Age (juvenile)	-0.01	0.11	-0.13	0.900
Model 2: Maze N= 31	Intercept	0.64	0.10	6.33	<0.001
	Number of errors[b]	-0.12	0.06	-1.97	0.059
	Sex (male)	-0.54	0.11	-4.89	<0.001
	Age (juvenile)	-0.18	0.12	-1.57	0.129

Reference categories for categorical predictors are indicated in brackets, SE: Standard error.

[a] Covariate was log transformed and afterwards z-transformed to a mean of= 0 and sd= 1; original mean of log(solving time) (sd)= 4.49 (0.96).

[b] Covariate was z-transformed to a mean of= 0 and sd=1; original mean (sd)= 13.61 (9.43).

Table S11: Results of the Cox proportional hazards model fitting the relationship between test performance in the food extraction task and survival

Predictor variable	**coeff**	**Exp(coeff)**	**SE(coeff)**	**z**	**P**
Solving time[a]	0.09	1.10	0.15	0.62	0.534
Sex (male)	-0.72	0.49	0.31	-2.35	0.019
Age (juvenile)	1.87	6.50	0.44	4.28	<0.001

Positive coefficients indicate a higher hazard (risk of death), i.e., a lower survival probability. Exponentially transformed coefficients are the hazard ratios and give the effect size on the hazard of predictor variables. Reference categories for categorical predictors are indicated in brackets, SE: Standard error, N= 64.

[a] Covariate was log transformed and afterwards z-transformed to a mean of= 0 and sd= 1; original mean of log(solving time) (sd)= 4.42 (0.93).

Table S12: Results of the Cox proportional hazards model fitting the relationship between test performance in the maze and survival

Predictor variable	**coeff**	**Exp(coeff)**	**SE(coeff)**	**z**	**P**
Number of errors[a]	-0.04	0.97	0.16	-0.23	0.824
Sex (male)	-0.75	0.47	0.31	-2.45	0.014
Age (juvenile)	1.69	5.41	0.46	3.63	<0.001

Positive coefficients indicate a higher hazard (risk of death), i.e., a lower survival probability. Exponentially transformed coefficients are the hazard ratios and give the effect size on the hazard of predictor variables. Reference categories for categorical predictors are indicated in brackets, SE: Standard error, N= 62.

[a] Covariate was z-transformed to a mean= 0 and sd= 1; original mean (sd)= 14.81 (9.27)

Chapter 3

Measuring fitness-related traits and their link to cognition in a wild primate

Franziska Hübner, Claudia Fichtel & Peter M. Kappeler

Author contributions:
FH, CF and PMK designed the study, FH conducted the fieldwork, analysed the data, and wrote the manuscript with input from CF and PMK.

Abstract

Research on animals' inter-individual differences in cognitive abilities and their fitness consequences is of growing interest recently. By now, various indicate that links between cognition and fitness in wild animals are complex and depend on various different factors, like a species' ecology, the specific measures investigated, but also on individual characteristics and non-cognitive correlated traits. Because most studies investigated only bivariate relationships, understanding trade-offs and complex interactions between a given cognitive ability and various fitness aspects is difficult. Therefore, we investigated different fitness-related traits in 86 wild grey mouse lemurs (*Microcebus murinus*) in order to link them with performance in four cognitive tasks addressing ecologically relevant cognitive abilities. We found that lemurs' physiological condition factor, summarizing measures of body condition, hematocrit and long-term cortisol levels, did not predict short-term survival and could therefore not be validated as a fitness proxy. In contrast to some other studies reporting links between cognitive performance and short-term survival, we found no correlation between cognitive performance and longevity in grey mouse lemurs, suggesting that the assessed cognitive abilities did not provide (net) benefits in survival. Our results further highlight the need to expand links between cognition and fitness within study species by investigating multiple cognitive abilities addressing fitness-related behaviours in different contexts and various fitness outcomes simultaneously. This will help to detect the complex relationships between cognition and fitness and broaden our understanding how cognition evolved.

Key words

Fitness proxies, cognitive performance, individual differences, survival, primate

Introduction

Unravelling the evolution of cognition is an exciting research area and it remains unclear how, why and when cognitive abilities evolved (Boogert et al., 2018). To understand how selection acts on cognition, studies have to link cognitive abilities with fitness outcomes in wild animals. This requires studying individual differences in cognitive performance, as opposed to focusing on species differences while ignoring individual variation, and linking this cognitive variation with measures of fitness. This endeavour is challenging in many ways: the difficulties in finding ecologically relevant and feasible cognitive tests and in quantifying reliable measures of cognitive abilities in wild animals has been discussed in detail (Rowe & Healy, 2014; Thornton et al., 2014; Morand-Ferron et al., 2016; Boogert et al., 2018). However, even if a wild animal's cognitive ability has been measured reliably while controlling for confounding factors like motivation, personality, experience or environmental influences (Morand-Ferron et al., 2016), identifying and quantifying meaningful fitness proxies imposes an additional challenge.

Evolutionary fitness can be defined as an individual's genetic contribution to future generations and correlates positively with individual differences in quality, i.e. traits associated with survival and reproduction (Wilson & Nussey, 2010). Commonly used proxies of individual fitness are therefore, apart from survival and reproductive success, correlated traits like individual growth and body condition indices (Stearns, 1989; Blums et al., 2005; Wilson & Nussey, 2010). Linking variation in cognitive abilities with variation in fitness is only relevant in wild populations, i.e., in environments to which organisms are adapted and where selection is operating (Ellegren & Sheldon, 2008). However, measuring key proxies of individual fitness, like reproductive success and survival, can be challenging for some taxa because large sample sizes need to be assembled over large temporal and spatial scales (Kingsolver et al., 2001; Morand-Ferron et al., 2016). Nevertheless, studying links between cognition and fitness across species and taxa is crucial to understand how cognition evolved (Kolm, 2014).

Because eggs, hatchlings and fledglings can be counted relatively easily and both putative parents can be often observed or even tested on cognitive abilities at the nest, many previous studies have examined the cognition-fitness link in birds (Cole et al., 2012; Cauchard et al., 2013; Preiszner et al., 2017; Wetzel, 2017; Ashton et al., 2018). Evidence for a link between cognitive performance and proxies of birds' mating or reproductive success has been mixed and differed between studies, species, sexes, reproductive success measures and cognitive measures (positive correlation: Keagy et al., 2009; Cole et al., 2012; Cauchard et al., 2013; Preiszner et al., 2017; negative correlation: Cole et al., 2012; no correlation: Preiszner et al., 2017; Isden et al., 2013; Wetzel, 2017). Direct reproductive success via parentage analysis was so far only analysed in rose bitterlings (*Rhodeus ocellatus*) in the laboratory and learning accuracy predicted reproductive success of males when using the sneaker role as a mating tactic (Smith et al., 2015).

Survival has been linked to cognitive performance, but existing studies have only addressed short-term survival, i.e. if subjects survived for 60 days (negative correlation for reverseal learning and correlation depending on body mass for associative learning: pheasants; *Phasianus colchicus:* Madden et al., 2018), until the following winter (no relationship: great tits; *Parus major*: Cole et al., 2012) or until the next breeding season (positive relationship for males, negative for females: African striped mice; *Rhabdomys pumilio*: Maille & Schradin, 2016). Moreover, foraging success has been used as a fitness proxy, in bumble bees (*Bombus terrestris*), which correlated positively with associative learning abilities in bee colonies (Raine & Chittka, 2008), but negatively on the individual bee level (Evans et al., 2017). Thus, different studies used different, mostly single fitness proxies, and their links with cognitive performance are heterogeneous across studies.

Since increased cognitive abilities also have costs, like higher energetic costs for the maintenance of additional neuronal tissue (Kawecki, 2010), assessing different fitness-associated parameters increases the chance to determine net benefits in fitness and to reveal trade-offs of increased cognitive abilities (Thornton et al., 2014, example in Cole et al., 2012). Therefore, in our study, we aimed at identifying several meaningful fitness proxies for a cognition-fitness link in a wild primate species. We assessed individual longevity as well as different measures of individual quality as fitness proxies, investigated variation in these measures, and linked them to cognitive performance in four cognitive tasks in a wild primate species.

Grey mouse lemurs (*Microcebus murinus*) are exceptionally well suited among primates for a study of several fitness proxies and their link with performance in different cognitive tests. The small (60g), solitary, nocturnal primates are ecological generalists which have a relatively large brain for their body size (Dammhahn & Kappeler, 2008b; MacLean et al., 2009). Grey mouse lemurs have a fast life history and reach sexual maturity within their first year of life and live on average two to three years in the wild (Kraus et al., 2008; Hämäläinen et al., 2014). They face multiple ecological challenges under which animals are likely to benefit from pronounced cognitive abilities (Roth et al., 2010) as they live in a seasonally changing environment, with a long dry season characterized by low food availability. Moreover, they are preyed upon by several types of predators and face one of the highest predation risks among primates (Scheumann et al., 2007; Rahlfs & Fichtel, 2010). Mouse lemurs can be captured easily with live traps (Eberle & Kappeler, 2002) and their high recapture probability allows to reliably estimate survival in the wild. Moreover, they adapt rapidly to short-term captivity and the presence of a human experimenter which allows testing the wild subjects during short-term captivity without long phases of habituation.

To assess different indices of physiological condition, we measured health and condition indicators that are likely to impact fitness**.** The different condition measures were combined into one composite factor of physiological condition, reflecting animals' condition at the end of the rainy season and we tested its link with short-term survival. This validation is important when using condition measures that reflect only short time spans as a proxy for fitness (Hõrak et al., 2002; Hatch & Smith, 2010; Barnett et al., 2015; Beehner & Bergman, 2017). First, we assessed body condition of individuals, which reflects variation in energetic state in small mammals (Schulte-Hostedde et al., 2005) and predicted survival or recruitment success in grey mouse lemurs (Rakotoniaina et al., 2017), as well as in several other species (Clutton-Brock et al., 1987; Tinbergen & Boerlijst, 1990; Blums et al., 2005; Bowers et al., 2014; reviewed in Barnett et al., 2015).

Secondly, we measured hematocrit, which is the volume percentage of erythrocytes in the blood and determines the ability to deliver oxygen to tissues, therefore functioning as an indicator of anaemia. It has been frequently used as an indicator of health and condition, for example in wild birds (Ots et al., 1998; Fair et al., 2007; Bowers et al., 2014). Above optimal hematocrit levels cause an increase in blood viscosity which hampers oxygen delivery and reduces cardiac efficiency, but also below optimal hematocrit levels lead to a reduced ability to carry oxygen, suggesting a nonlinear relationship between hematocrit and individual fitness, which was found in bird hatchlings (Birchard, 1997; Bowers et al., 2014).

Thirdly, glucocorticoids, i.e., cortisol, served as a further physiological indicator of relative condition and health (Bonier et al., 2009; Walker et al., 2005). Glucocorticoids play a key role in mediating various physiological processes, and glucocorticoid levels are commonly interpreted as an indicator of stress or allostatic load (Korte et al., 2005). Therefore, the Cort-Fitness hypothesis predicts that high levels of baseline cortisol indicate an individual in worse condition (Bonier et al., 2009). However, this hypothesis has been questioned lately because elevated glucocorticoid levels correspond to mobilized energy under environmental challenges, therefore providing fitness benefits in this situation (Bonier et al., 2009; Beehner

& Bergman, 2017). To noninvasively assess long-term basal cortisol levels in the wild mouse lemurs, we measured hair cortisol concentrations (HCC) (Fourie et al., 2016). HCC reflects average cortisol levels accumulated over time periods of up to several months as cortisol is incorporated into the growing hair shaft (Stalder & Kirschbaum, 2012). In addition, hair cortisol levels of grey mouse lemurs in our study area have been shown to predict survival (Rakotoniaina et al., 2017).

Longevity served as the most direct fitness proxy as we could follow most of our subjects for their entire life. For juveniles, surviving the first dry season determines the chance to reproduce at all. Females start to reproduce in their first year and continue to have offspring once a year, thus female reproductive skew is negligible and longevity should correlate with lifetime reproductive success (Eberle & Kappeler, 2004a; Zimmermann et al., 2016). For males, surviving at least until the second year appears crucial as male reproductive success is determined by their body mass during the mating season, thus outcompeting adult competitors is difficult for males in their first year. In a prior study, we linked survival until the end of 2017 with two of the cognitive performance measures, but did not find that cognition predicted animals' survival during this period.

We linked subjects' cognitive abilities in four cognitive tests addressing different ecologically relevant cognitive abilities with these fitness proxies. We tested associative motor learning during repeated and spontaneous innovative problem solving (Griffin, 2016; Griffin & Guez, 2014), which is of general ecological relevance as innovations allow animals to exploit new resources or to use existing resources more efficiently (Griffin & Guez, 2014; Reader & Laland, 2003). Spatial learning and memory, i.e. the ability to remember the location of important resources, is crucial for mouse lemurs because they rely on sparsely distributed but predictable food resources during long and harsh dry seasons (Dammhahn & Kappeler, 2008a; Lührs et al., 2009). Finally, inhibitory control, i.e. the ability to inhibit prepotent responses, is involved in various decision-making processes in asocial but also social contexts (Hauser, 1999; Amici et al. 2008; MacLean et al., 2014). Cognitive performance in these tasks was shown to differ between individuals, to not be systematically influenced by non-cognitive factors like personality and motivation and could not be summarized within one general factor of cognitive performance (Huebner et al. submitted).

Methods

Study population and general procedure

The study was conducted at Kirindy Forest (CNFEREF), a dry deciduous forest in central Western Madagascar, at the research station of the German Primate Center. Grey mouse lemurs inhabiting a 10ha study area were regularly captured between the beginning of the dry season in April and the beginning of the rainy season in November across four years (2015 – 2018). Animals were captured with Sherman live traps and marked individually with subdermal microtransponders (Trovan Euro I.D., Frechen, Germany) while being sedated with 0.01ml Ketanest 100 (see Rensing, 1999). Standard morphometric measures were taken and individuals were sexed and aged (juveniles: less than 10 months old) based on their body mass and size collected at the time of first capture (Dammhahn & Kappeler, 2008a).

In total, 97 individuals (44 females, 53 males, 63 juveniles, 34 adults) participated in the cognitive experiments of this study. For the experimental testing, animals were singly housed in the research station in 1m^3 cages and kept for up to three consecutive nights before they were released again in the evening to their specific site of capture. Tests were conducted at night under dim red light. Subjects participated voluntarily in the test and were rewarded with small pieces of banana in all cognitive tests and obtained a 1.5cm banana piece per night after the testing, while water was provided *ad libitum* (for more details on the housing and testing conditions see Huebner et al., 2018).

Subjects were tested with the cognitive test battery in a determined order (as presented below). All test sessions were video-recorded and analysed with the software BORIS (Friard & Gamba, 2016).

Measures of fitness-related traits

Body mass index

To estimate body condition, we calculated a body mass index (BMI) by dividing body mass (g) by bizygomatic breadth (mm), the latter being a reliable measure of body size in grey mouse lemurs (Rasoloarison et al., 2000).

Hematocrit

We collected blood samples in heparinized microhematocrit capillary tubes during regular capture events when animals were sedated for the general handling procedure (see above). Blood was taken from the femoral vein and directly centrifuged at 10 000 x g for 5 minutes. Hematocrit was measured as the percentage of packed red blood cells relative to the whole blood volume. Additional to the original hematocrit values, we calculated for each month subjects' absolute deviation from the population mean as a proxy for optimal hematocrit values. In total, hematocrit was obtained for N= 182 individuals in the months April to July.

Cortisol

Hair samples for cortisol analysis were taken in the months April and May, thus reflecting cortisol levels during the rainy season. Samples were taken from the animals' dorso-caudal region using a pet grooming clipper (Aesculap Isis GT 420) to cut the hair as close as possible to skin. As the mouse lemurs' hair needs about two months to fully regrow from these shavings (personal observation), we are confident that the obtained samples reflect hair cortisol concentrations of a time span of at minimum two months before sampling. The analysis of hair samples was carried out at the University of Dresden (Kirschbaum lab in Germany) and followed an adjusted standard protocol described by Gao et al. (2013) and Rakotoniaina et al. (2017).

Short-term survival and longevity

We used short-term survival as a fitness proxy for the validation of the physiological condition factor. To this end, we assessed subjects' survival as a binary outcome in the year of their cognitive tests (i.e. typical the year of first encounter within the study period) between April and October. This period reflects the biggest part of the long, harsh dry season in Kirindy forest, during which food availability is low and females hibernate for several months, while males use shorter torpor bouts to save energy before the mating season starts in October (Schmid & Kappeler, 1998; Schmid, 1999; Rasoazanabary, 2006; Vuarin et al., 2013). To be able to hibernate, individuals have to accumulate body mass during the rainy season and juveniles have to complete growth at about six months of age. Thus, subjects' physiological condition factor reflecting condition at the end of the rainy season is likely to predict survival during this period.

We estimated long-term survival until November 2018 by determining an individual's number of days alive between birth and the date of last capture, truncating the study period in November 2018. We set birth dates for all individuals at the modal birth date 1 January of the year of first capture for juveniles and one year earlier for subjects firstly captured as adults (see Eberle & Kappeler, 2004b). As natal dispersals occur within the first year of life (Schliehe-Diecks et al., 2012) and the probability of not capturing a natal individual within the first year of life is presumably extremely small, we were also able to assign birth dates to adults. To define death operationally for individuals not recaptured for longer periods, we determined the 95th percentile of the frequency distribution of 10 936 inter-capture intervals recorded between 1995 and 2017 as a cut-off point and considered individuals dead if they were not recaptured within 161 days before 1 November 2018. In total, we could estimate survival for 86 subjects that also participated in the cognitive tasks, excluding 11 juvenile males for which we could not rule out dispersal from the study area in their first year of age. Only 7 individuals were still alive at the date of censoring.

Cognitive test battery

Food extraction task

In this novel motor task, subjects had to solve a novel problem repeatedly by removing sliding covers on six wells (5 x 4.5cm) of a small task box (Fig. S1a), each containing a food reward. A small banana piece on top of the task box served as an initial incentive to interact with it. During 20 minutes of testing, we for subjects that opened at least two lids their solving time (s), i.e. the mean time they spent per successful opening after having opened the first lid. This measure is repeatable and addresses a subject's efficiency in learning to apply the novel motor action (Huebner et al., 2018, Huebner et al. submitted).

String-pulling task

In this task we addressed subjects' causal understanding and spontaneous problem solving. Banana attached to a string of 20cm length was positioned outside of the test cage, with the other end within reach of subjects through the cage wire (Fig. S1b). During 20 minutes of testing, we recorded subjects' latencies from the first attention to the reward until the successful pulling of the string. If subjects failed to pull the string and did not obtain the reward, we assigned maximal latencies.

Spatial learning

During the spatial learning task, subjects' ability to remember the position of a food reward in a plus maze was tested. The maze consisted of four wooden arms (40x17cm; Fig. S1c) with attached boxes (20x17cm) at each arm's end. One of the boxes served as the starting point from where subjects were released into the maze, and either the arm to the left or the right served as goal box that contained a small banana reward. To control for olfactory cues, large banana pieces were placed out of reach at the end of each maze arm, masking the smell of the actual reward. In order to avoid subjects from using own odour trails, the maze was cleaned with 70% ethanol after every third trial. Before testing, subjects had to pass a familiarization trial where they had to find rewards in all three arms of the maze. Animals were then tested in 15 test trials, each trial started with the release of the subject from the start box and ended with it obtaining the reward in the goal box. We recorded the number of errors made, i.e. the number of times animals entered an unrewarded maze arm, until reaching the learning criterion of finding the reward without errors in three consecutive trials or until the end of testing for animals that never met criterion. We graded the error scores to differentiate between the different levels of entering wrong arms. We assigned error scores of 1 if a subject entered the box at the end of an unrewarded arm, a score of 0.5 if a subject entered the arm but not the box at the end of the arm and a score of 0.25 when it entered a wrong arm with only part of the body.

Inhibitory control

We tested individuals' inhibitory control by assessing their ability to inhibit an ineffective prepotent response towards a food reward (MacLean et al., 2014). This detour-reaching task consisted of an open-ended transparent cylinder (20cm length, 6cm diameter, Fig. S1d, right cylinder) containing a clearly visible food reward in the centre. To control for odour cues that subjects might follow to retrieve the reward, small holes were made in the centre of the cylinder and the cylinder was cleaned on every fourth trial with 70% ethanol. For each trial, subjects were attracted with a small reward to one corner of their test platform before the cylinder was placed in the test cage so that subjects started to reach it from a central position at about 40cm distance. In a familiarization phase, subjects had to retrieve a food reward out of an opaque cylinder (Fig. S1d, left cylinder) in five consecutive trials in order to start the test phase with the transparent cylinder. For the 10 trials of the test phase, we scored the number of erroneous trials subjects made when trying to first directly reach the reward before detouring to the open ends of the cylinder.

Statistical analysis

To investigate the effect of individual characteristics and different extrinsic factors on variation in the measures of physiological condition, we fitted linear mixed models (LMM) using R package lme4 (Bates et al., 2015) with hematocrit or cortisol as a response, sex and age category as a fixed factors and individual ID as a random factor. To control for the effect of different sampling years, year was included as a further fixed effect. For hematocrit, we calculated two separate models, one with absolute hematocrit values as a response and one with the absolute deviation from mean hematocrit values as a proxy for optimal hematocrit as response. Moreover, we included test month as a further fixed factor in the models to investigate seasonal changes with increasing test months. Prior analysis, cortisol and deviations from hematocrit means were log-transformed.

In order to combine the three different measures of physiological condition into one factor, we first tested correlations between deviations from mean hematocrit, BMI and cortisol using spearman rank correlations. Then we performed a principal component analysis (PCA) with an unrotated factor solution and extracted the first principal component (PC1) with an eigenvalue >1 as the combined physiological condition factor. We tested sampling adequacy of the correlation matrix used in the PCA with the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and the Bartlett’s test of sphericity, which was considered appropriate with a KMO> 0.5 and Bartlett’s test of P< 0.05 (Budaev, 2010).

To validate this factor of physiological condition, we examined its relation to short-term survival. To this end, we implemented survival (yes/no) as a response in a generalized linear model (GLM) with binomial error structure and logit link function, and the condition factor (PC1), sex, age and year as predictors. Because we cannot exclude dispersal for juvenile males, we compared model results with a second model, fitted with a reduced dataset with only females, thus lacking the sex predictor.

To test the effect of cognitive performance on longevity, we fitted linear models (LM) with the (log-transformed) number of days an animal lived as response. We could not fit Cox proportional hazards models for this survival analysis as the assumption of proportional hazards was violated. Therefore, we fitted always two sets of models, one including all tested individuals and one lacking the individuals that were still alive at the end of observation, resembling censored observations in a Cox model (N= 7, reflecting not only the oldest subjects of the dataset). In these models, we included the respective measure of cognitive performance as a covariate and sex and age category as control factor first with performance measures in separate models to increase sample size. In the final models, we included all four measures of cognitive performance together in one model for subjects that were tested in all cognitive tasks, resulting in a reduced sample size.

For all models, we z-transformed covariates to a mean of 0 and a standard deviation of 1 prior to fitting to facilitate interpretation of predictor estimates (Schielzeth, 2010), except for the physiological condition factor in the short-term survival models. In all models, we first tested the interactions between sex and age, sex or age and physiological condition, or sex or age and cognitive performance, respectively, but removed these interactions from the models if they were not significant but the respective full null model comparison with the interaction and main effect removed was significant. We always checked the model assumptions "absence of collinearity" using Variance Inflation Factors (Fox & Monette, 1992; package car in R: Fox & Weisberg, 2011) and "absence of influential observations" using dfbetas in all models. We controlled for the effect of potential outliers/ influential cases by comparing model results fitted with and without these observations and retained the complete dataset if results did not change. For the LMs testing the effect of all cognitive performance measures on longevity (see model S7a and S7b) we had to exclude one outlier as it changed model results significantly. For LMM and LM, we visually checked normally distributed and homogenous residuals. We always tested our full model against a null model containing the intercept only or just control predictors with an F-test for LM and a likelihood ratio test for LMM and GLM. For LMMs, p-values for individual predictors were obtained using likelihood ratio tests comparing the full with respective null model (Barr et al., 2013; function drop1 in R). All analyses were conducted in R, v. 3.4.2 (R Core Team, 2017) and the level of significance was set at 0.05.

Results

Indices of physiological condition

Hematocrit

We did not find that hematocrit levels differed between the sexes, but hematocrit significantly increased over time, i.e., across the dry season, and was higher in adults (full null model comparison: X^2= 15.384, df= 3, P= 0.002, Table 1). However, we found a sex difference in animals' deviation from mean hematocrit as males had significantly higher absolute deviations from the estimated optimal hematocrit (full null model comparison: X^2= 11.079, df= 3, P= 0.011, Table 2), but age class and month did not predict deviations from mean hematocrit (Table 2).

Table 1: Results of the linear mixed model testing the effect of predictors on hematocrit

Term	**Estimate**	**SE**	**CI_{lower}**	**CI_{upper}**	**X^2**	**df**	**P**
Intercept	48.215	0.473	47.283	49.147	[a]	[a]	[a]
Month*	0.784	0.227	0.337	1.231	11.583	1	<0.001
Age category (juvenile)	-1.281	0.449	-2.165	-0.397	8.003	1	<0.01
Sex (male)	0.089	0.438	-0.777	0.951	0.041	1	0.839
Year (2016)[b]	-0.059	0.490	-1.023	0.906	0.830	2	0.661
Year (2017)[b]	-0.488	0.568	-1.607	0.630			

Given are the estimated coefficients for each predictor and associated standard errors (SE), lower and upper limits of the 95% confidence intervals and the likelihood ratios (X^2) with degrees of freedom and P values. N= 272 for 182 individuals.

a) Not shown due to a very limited interpretation (testing the null hypothesis that the estimate, here intercept, is equal to zero).

b) Indicated test was obtained from a likelihood ratio test comparing the full model with a reduced model lacking the year predictor.

* Covariate was z-transformed, mean and sd of the original values were 5.08 and 1.09, respectively.

Table 2: Results of the linear mixed model results testing the effect of predictors on absolute deviation from hematocrit mean

Term	**Estimate**	**SE**	**CI$_{lower}$**	**CI$_{upper}$**	**X^2**	**df**	**P**
Intercept	0.192	0.133	-0.068	0.453	[a]	[a]	[a]
Month*	0.065	0.071	-0.075	0.205	0.829	1	0.363
Age category (juvenile)	0.046	0.130	-0.209	0.301	0.127	1	0.722
Sex (male)	0.368	0.122	0.128	0.608	8.951	1	< 0.01
Year (2016)[b]	0.043	0.148	-0.248	0.334	8.708	2	0.013
Year (2017)[b]	0.474	0.169	0.141	0.807			

Given are the estimated coefficients for each predictor and associated standard errors (SE), lower and upper limits of the 95% confidence intervals and the likelihood ratios (X^2) with degrees of freedom and *P* values. N= 271 for 181 individuals.

a) Not shown due to a very limited interpretation.

b) Indicated test was obtained from a likelihood ratio test comparing the full model with a reduced model lacking the year predictor.

* Covariate was z-transformed, mean and sd of the original values were 5.08 and 1.09, respectively.

Hair cortisol concentration

We found a significant interaction between sex and age class predicting hair cortisol concentrations during the rainy season. In adults, cortisol concentrations differed significantly between the sexes and were higher in females, whereas for juveniles, we did not find a significant sex difference (full null model comparison: X^2= 50.7, df= 3, P< 0.001, Table 3, Fig. S2). For males, cortisol concentration was significantly increased in juveniles compared to adults, whereas in females, this difference was not significant (Table 3, Fig. S2).

Table 3: Results of the linear mixed model results testing the effect of predictors on hair cortisol concentration in the rainy season

Term	**Estimate**	**SE**	**CI*lower***	**CI*upper***	**X^2**	**df**	**P**
Intercept	2.570	0.140	2.293	2.848	[a]	[a]	[a]
Age category (juvenile)	0.343	0.189	0.031	0.717	[a]	[a]	[a]
Sex (male)	-0.747	0.160	-1.064	-0.433	[a]	[a]	[a]
Age.cat *sex	0.962	0.245	0.475	1.446	14.396	1	< 0.001
Year (2016)[b]	-0.112	0.132	-0.372	0.148	4.250	2	0.119
Year (2017)[b]	-0.376	0.181	-0.733	-0.019			

Given are the estimated coefficients for each predictor and associated standard errors (SE), lower and upper limits of the 95% confidence intervals and the likelihood ratios (X^2) with degrees of freedom and *P* values. N= 139 for 97 individuals.

a) Not shown due to a very limited interpretation.

b) Indicated test was obtained from a likelihood ratio test comparing the full model with a reduced model lacking the year predictor.

Physiological condition factor

Individuals differed in the three physical condition proxies at the end of the rainy season, i.e.: BMI, deviation from mean hematocrit and hair cortisol concentration (Fig. S3). Individuals' BMI and cortisol levels correlated significantly negatively (Spearman rank correlation, rho= -0.37, P< 0.001, N= 80), but deviations from hematocrit means did neither correlate with BMI (rho= -0.11, P= 0.34, N= 80) nor cortisol levels (rho= 0.07, P= 0.55, N= 80). Subjects' overall physiological condition at the end of the rainy season could be summarized into a first principal component that contributed to 45% of the total variance (PCA; N= 80, Table 4; KMO= 0.52, Bartlett's test: P< 0.05). Subjects' BMI loaded positively, whereas cortisol and deviation from mean hematocrit levels loaded negatively on the first principal component, resulting in a condition factor with high values for animals in good physiological condition and low values for individuals in poorer condition.

Table 4: Results of the principal component analysis

Proxy of physiological condition	**PC1**	**PC2**	**PC3**
BMI	0.671	-0.246	0.670
Hair cortisol concentration	-0.680	0.172	0.712
Deviation hematocrit mean	-0.295	-0.954	-0.052
Eigenvalue	1.160	0.984	0.827
% variance explained	0.45	0.32	0.23

Is the general physiological condition factor a good fitness proxy?

The physiological condition factor did not predict short-term survival across the dry season (GLM with all subjects: N= 80, full null model comparison: X^2= 1.51, df= 4, P= 0.83, Table S1, Figure 1; GLM for females only: N= 36, full null model comparison: X^2= 1.36, df= 2, P= 0.51, Table S2). Therefore, we did not further use the condition factor as a proxy for fitness and did not test the link between cognitive performance and physiological condition.

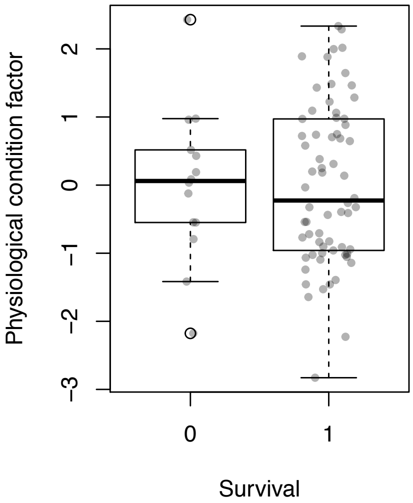

Figure 1: The physiological condition factor of individuals that survived (1) and did not survive (0) until the beginning of the breeding season (short-term survival). Individual data points are jittered to increase visibility.

The link between cognitive performance and fitness proxies

Does cognitive performance predict longevity?

We did not find that cognitive performance in any of the four tests predicted longevity, i.e. the number of days subjects lived (LMs: Food extraction task: solving time: full null model comparison: $F_{-1,63}$= 1.09, P= 0.3, N= 66, Table S3a; String pulling latency: full null model comparison: $F_{-1,83}$= 0.87, P= 0.35, N= 86, Table S4a ; Maze: errors until criterion: full null model comparison: $F_{-1,62}$= 0.44, P= 0.51, N= 65, Table S5a; Detour-reaching performance: Full null model comparison: $F_{-1,55}$= 0.07, P= 0.79, N= 58, Table S6a). Also when we implemented all cognitive performance predictors together in one model, we did not find a significant effect of cognitive performance on the number of days an animal lived (LM: full null model comparison: $F_{-4,37}$= 1.84, P= 0.14, N= 40, Table S7a). Model results with the reduced data set lacking animals that were still alive did not differ meaningfully (Tables S3b to S7b).

Discussion

Indices of physiological condition - natural variation and validity as fitness proxy

We investigated variation in the measures of physiological condition that could be summarized into one physiological condition factor explaining a moderate amount of the total variance. However, we could not validate this condition factor to be a meaningful fitness proxy as it did not predict short-term survival. Several factors may explain the lack of this relationship.

Firstly, variation in the three measures that we combined to describe physiological condition can be partially explained by extrinsic and intrinsic factors like season and individuals' reproductive and developmental stages. Hematocrit levels in mouse lemurs were higher in adults compared to juveniles because growing individuals still increasingly produce red blood cells (Fair et al., 2007). Furthermore, we found that hematocrit levels increased with advancing months of sampling, which might be explained by corresponding reductions in humidity and food availability (Fair et al., 2007). Therefore, we used only measures at the end of the rainy season for the combined physiological condition factor. Deviations from the approximated optimal hematocrit level were low, albeit higher in males, which is difficult to explain and has not been reported in other studies. Only two individuals had haematocrit values of <40%, which might reflect critically low values of anaemic animals (for example <35% in birds: Campbell, 1994 and 36-40% in humans: Billett, 1990). In contrast, in house wren (*Troglodytes aedon*) nestlings hematocrit ranged between 22 to 66% and predicted recruitment success and longevity (Bowers et al., 2014). In birds, blood samples of nestlings can be taken directly at the nest, so that even weak individuals will not be missed. However, in our study we could only sample active mouse lemurs and may have missed individuals in a critically weak physiological condition. In this context it is worth pointing out that we did not capture any mouse lemurs with extremely low body mass in 2357 capture events between 2015 and 2017. Thus, hematocrit measures may have been biased towards animals in good condition.

Hair cortisol concentrations during the rainy season were higher in adult females and juveniles. Glucocorticoids are secreted during energy-demanding situations or periods in response to extrinsic (e.g. ecological changes) or intrinsic challenges (e.g. different developmental and reproductive stages), therefore constituting an adaptive response, the “predictive homeostasis” (Romero et al., 2009; Beehner & Bergman, 2017). Therefore, higher cortisol concentrations in adult females, for which the measured period included lactation, and higher HCC in growing juveniles can be expected and were also found in other studies with primates (e.g. Fourie et al., 2016; Rakotoniaina et al., 2016). Thus, ecological challenges or intrinsic factors that themselves are likely to impact fitness have to be controlled for when linking glucocorticoids to fitness (Johnstone et al., 2012; Beehner & Bergman, 2017). The two previous studies that reported a link between glucocorticoid levels and survival in primates did not control for these potentially confounding factors, however. In ring-tailed lemurs (*Lemur catta*) mortality across two years was higher in individuals with above average glucocorticoids (Pride, 2005). Similarly, high levels of HCC were associated with reduced survival in grey mouse lemurs (Rakotoniaina et al., 2017), but both studies did not account for age effects. Thus, variation in our chosen physiological condition factor may not reflect individual health or quality, but rather differences in energetically costly conditions like reproductive and developmental stages, and does therefore not predict survival when controlling for these factors.

Secondly, surviving the harsh dry season is not only determined by physiological condition but also by predation, which is probably the leading cause of death in this species (Goodman et al., 1993) and could mask direct effects of physiological condition on mortality.

However, poor physiological condition might not only increase individuals' susceptibility to disease (Coop & Kyriazakis, 1999) but also to predation (Murray, 2002), so that interactions between condition and predation may affect survival. This is in line with the finding that body mass was lower in the season preceding death in the same mouse lemur population (Hämäläinen et al., 2014; Rakotoniaina et al., 2017). Yet, the sample size in the current study might be too low to detect such a condition-dependent mortality.

The link between cognition and longevity

Variation in survival is affected by various extrinsic factors, like resource availability, disease, competition and predation, and intrinsic factors, like age, sex and general physiological condition, as well as their various interactions (Jorgenson et al., 1997; Lindström, 1999; Farand et al., 2002; Blums et al., 2005; Ozgul et al., 2006; Wilson & Nussey, 2010). Cognitive abilities can help individuals to survive by facilitating adaptations to changing environmental conditions, maintaining a healthy body condition by more efficient foraging, enhanced spatial memory of resources and better predator avoidance (e.g. Pravosudov & Clayton, 2002; Sol et al., 2002; Sol et al., 2007; Raine & Chittka, 2008; Roth et al., 2010; Maille & Schradin, 2016).

We tested grey mouse lemurs in different ecologically relevant cognitive abilities, including innovative problem solving and subsequent learning of novel motor patterns, spatial learning and inhibitory control. However, performance in none of the cognitive tasks predicted longevity. Thus, the cognitive abilities we tested here do not seem to provide grey mouse lemurs with survival advantages. The tests we used mainly addressed cognitive abilities that are likely to improve foraging success, which contributes to body condition, e.g. through innovative foraging or better spatial memory of dispersed food items. Indeed, for subjects' ability to quickly and efficiently learn the new motor action in the food extraction task, we could demonstrate that efficient solvers were better in maintaining body condition across the dry season (Huebner et al., 2018); yet, this advantage did not translate into better survival.

However, as mentioned above, grey mouse lemurs' survival is determined by strong predation by several different predators, including owls, snakes and carnivores (Goodman et al., 1993; Fichtel, 2016), and physiological condition did not predict short-term survival in our population, suggesting that cognitive abilities involved in predator avoidance rather than in foraging are important for mouse lemurs' survival. Thus, if the addressed cognitive abilities in our study do not provide clear benefits in avoiding, detecting and escaping these predators, this could explain the difficulty to detect links with longevity. In a study that directly addressed antipredator behaviours in African striped mice, links between test performance and subsequent short-term survival were detected (Maille & Schradin, 2016). More specifically, the response to a predator stimuli and the ability to remember shelter locations in a maze using a predator dummy as motivative incentive predicted survival until the beginning of the breeding season in the wild (Maille & Schradin, 2016). However, the directions of correlations differed between the sexes in striped mice, indicating that even when linking directly predation-related behaviours with survival, these links are not always straightforward and as predicted.

Testing cognitive abilities related to predator avoidance is difficult in mouse lemurs, however. As grey mouse lemurs are nocturnal, vigilance and subsequent fleeing to a distant shelter are not effective (Fichtel, 2016). After having detected a predator, grey mouse lemurs tend to freeze and behave cryptically (Rahlfs & Fichtel, 2010), a behaviour that is difficult to address in a cognitive task. We assumed that inhibitory control, i.e., the ability to inhibit the prepotent response of continuing foraging in the presence of a predator (van Horik et al., 2018) could play an important role in the lemurs' freezing behaviour. However, we also did not find a correlation between inhibitory control performance and survival, which might indicate that predator detection and predator-sensitive foraging are more crucial in this context (Dammhahn & Almeling, 2012; Fichtel, 2016).

In pheasants, chicks that were faster to reverse a learned association were less likely to survive for 60 days after being released into semi-wild conditions, which might indicate that also putative "better" cognitive abilities can be associated with maladaptive, costly fitness outcomes (Madden et al., 2018). In grey mouse lemurs, we also did not find a significant negative correlation between cognitive performance and longevity, even when we included all tasks together into one model. The cognitive abilities we addressed here do apparently not impact survival in either way. However, as large sample sizes are required to protect against Type II error in selection analyses (Kingsolver et al., 2001; Dingemanse & Reale, 2005; Morand-Ferron et al., 2016), we cannot exclude the possibility that a lack of statistical power contributed to the non-significant results, especially when investigating longevity in the few subjects that participated in all cognitive tasks.

In great tits, problem-solving ability in a lever-removing task did not predict survival, but birds that were able to solve a novel problem had larger clutches compared to non-solvers, probably because solvers were more efficient in exploiting their environment (Cole et al., 2012). Because reproductive success is highly dependent on the physiological condition of animals and is often traded off against survival (Roff, 1992, 2002; Stearns, 1992), investigating the link between cognitive performance and reproductive success will help to complete the links between cognition and fitness and investigate the adaptive value of cognitive abilities in grey mouse lemurs in the future. Unfortunately, we could not include reproductive success via parentage analysis in this study because samples of an entire year were destroyed, presumably due to storage problems in Madagascar, resulting in missing DNA for 38 individuals of the study population.

Critically, cognition is not a unitary trait and various cognitive processes are involved in shaping a given behavioural outcome. Also, cognitive processes are involved in various different contexts and what is beneficial in one situation might be disadvantageous in another, therefore not necessarily resulting in net benefits of fitness (Rowe & Healy, 2014; Thornton et al., 2014). Furthermore, cognitive abilities themselves are associated with fitness costs like increased energetic costs for neuronal tissue and are therefore also under selection (Kotrschal et al., 2013; Morand-Ferron et al., 2016). Therefore, direct links between cognitive test performance in artificial tasks and fitness outcomes can be very difficult to detect or might not even be expected (Rowe & Healy, 2014; Ten Cate, 2014). Thus, only detailed studies of cognitive abilities addressing different ecologically relevant behaviours in different contexts and their link with various fitness outcomes will help to understand these trade-offs, selection pressures and various fitness consequences and can finally lead to a better understanding how and why cognition evolved.

In conclusion, this study revealed the challenges in assessing meaningful fitness-related traits in a wild primate and highlights the importance of validating these measures before using them as a proxy for fitness. Our results on the cognition-fitness link demonstrate that better cognitive abilities do not necessarily translate into better survival, especially when testing links between cognitive performance and survival in the long-term, i.e. longevity of individuals. As predicted, links between cognition and fitness are complex and studying correlations between multiple cognitive and fitness traits will allow to detect the underlying selective pressures, diverse covariations and fitness trade-offs in the future.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (awarded to P.M.K.: Ka 1082/33-1). We are very grateful to Bruno Tsiverimana, Léonard Razamanantsoa and all other members of the Kirindy research station team for their support in the field. Moreover, we thank Roger Mundry for a brief, helpful discussion about the statistical analysis of the longevity data.

Supplemental material

Cognitive test battery

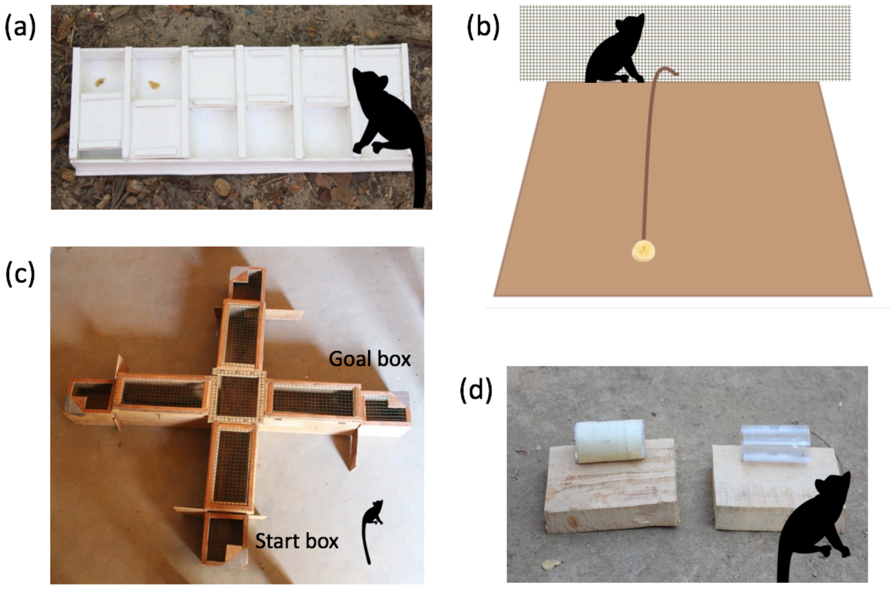

Figure S1: The cognitive test apparatuses.

(a) Food extraction task, (b) String-pulling task, (c) Plus maze, (d) Detour-reaching task

Approximate proportional size of a mouse lemur indicated by silhouette.

Measures of physiological condition

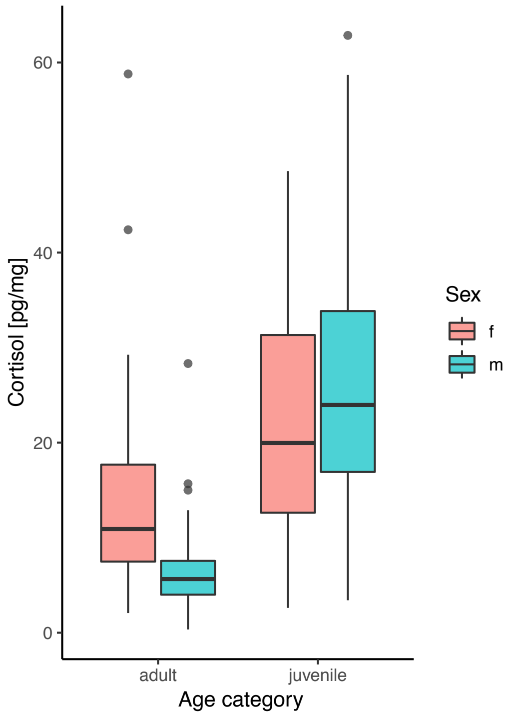

Figure S2: Hair cortisol concentration reflecting the rainy season, depicted for males and females in the two age classes.

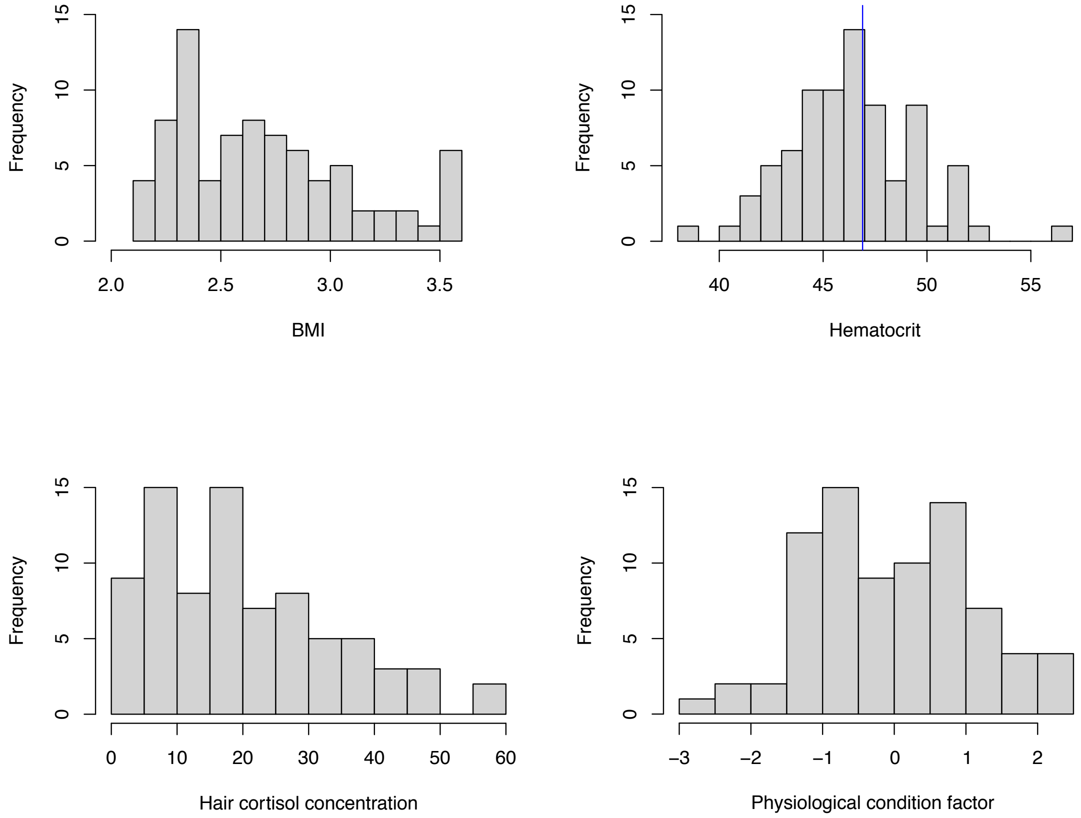

Figure S3: Histograms of the different proxies of physiological condition and the composite physiological condition factor. N= 80. For hematocrit, the population mean at 46.9% is depicted in blue. Hematocrit values are given in percent, hair cortisol concentration is given in pg/mg.

Is the general physiological condition factor a good proxy for individuals’ fitness?

Table S1: Result of the Generalized Linear Model (GLM) fitting the effect of the general physiological condition factor on survival probability until the breeding season (survival y/n) for all individuals

Predictor variable	Estimate	SE	z	P
Intercept	2.032	0.778	2.61	<0.01
Condition factor	-0.275	0.384	-0.72	0.473
Sex (male)	0.245	0.617	0.40	0.691
Age (juvenile)	-1.042	0.937	-1.11	0.266
Year (2016)	0.208	0.636	0.33	0.743

Reference categories for categorical predictors are indicated in brackets, SE: Standard error, N= 80.

Table S2: Result of the Generalized Linear Model (GLM) fitting the effect of the general physiological condition factor on survival probability until the breeding season (survival y/n) for females only

Predictor variable	Estimate	SE	z	P
Intercept	1.857	0.883	2.10	0.035
Condition factor	0.152	0.464	0.33	0.743
Age (juvenile)	-0.759	1.142	-0.67	0.506

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 36.

To reduce model complexity, we excluded year as a covariate as it did not significantly affect survival in the model with all subjects.

Does cognitive performance predict longevity?

Table S3a: Food extraction task: results of the linear model testing the effect of cognitive performance (solving time) on individuals’ survival in days

Predictor variable	Estimate	SE	t	P
Intercept	6.816	0.112	60.94	<0.001
FE solving time[a]	-0.062	0.059	-1.046	0.3
Sex (male)	0.347	0.118	2.951	<0.01
Age (juvenile)	-0.914	0.125	-7.338	<0.001

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 66.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 138.6 (168.4) sec.

Table S4a: String-pulling task: results of the linear model testing the effect of cognitive performance (success latency) on individuals' survival in days

Predictor variable	Estimate	SE	t	P
Intercept	6.841	0.089	76.935	<0.001
Success latency[a]	0.048	0.051	0.931	0.355
Sex (male)	0.309	0.1	3.088	<0.01
Age (juvenile)	-0.91	0.103	-8.836	<0.001

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 86.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 281.6 (364) sec.

Table S5a: Maze: results of the linear model testing the effect of cognitive performance (number of errors until criterion) on individuals' survival in days

Predictor variable	Estimate	SE	t	P
Intercept	6.66	0.115	57.772	<0.001
Errors until criterion[a]	0.036	0.055	0.664	0.509
Sex (male)	0.367	0.108	3.398	0.001
Age (juvenile)	-0.746	0.119	-6.26	<0.001

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 65.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 14.5 (9.2).

Table S6a: Detour-reaching task: results of the linear model testing the effect of cognitive performance (% correct responses) on individuals' survival in days

Predictor variable	Estimate	SE	t	P
Intercept	6.677	0.141	47.432	<0.001
% correct[a]	-0.017	0.064	-0.264	0.793
Sex (male)	0.354	0.132	2.683	0.01
Age (juvenile)	-0.773	0.134	-5.776	<0.001

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 58.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 69.1 (22.5) %.

Table S7a: Results of the linear model testing the effects of cognitive performance of all cognitive tests on individuals' survival in days

Predictor variable	**Estimate**	**SE**	**t**	**P**
Intercept	6.681	0.17	39.276	<0.001
Food extraction: solving time[a]	-0.202	0.14	-1.441	0.159
String pulling: latency success[a]	0.268	0.102	2.617	0.013
Maze: errors until criterion[a]	-0.004	0.077	-0.049	0.961
Detour reaching: % correct[a]	-0.047	0.079	-0.596	0.556
Sex (male)	0.308	0.158	1.952	0.06
Age (juvenile)	-0.716	0.186	-3.846	0.001

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 40.

Please note that outlier individual with extreme high solving time in the food extraction task was excluded from the dataset, as it changed model results significantly.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1.

Longevity models with the reduced dataset, lacking animals that are still alive

Table S3b: Food extraction task with reduced dataset: results of the linear model testing the effect of cognitive performance (solving time) on individuals' survival in days

Predictor variable	**Estimate**	**SE**	**t**	**P**
Intercept	6.702	0.119	56.395	<0.001
FE solving time[a]	-0.056	0.057	-0.992	0.326
Sex (male)	0.403	0.121	3.327	<0.01
Age (juvenile)	-0.85	0.135	-6.286	<0.001

Reduced dataset without animals that are still alive.

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 59.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 134.8 (176.2) sec.

Table S4b: String-pulling task with reduced dataset: results of the linear model testing the effect of cognitive performance (success latency) on individuals' survival in days

Predictor variable	Estimate	SE	t	P
Intercept	6.79	0.09	75.729	<0.001
Success latency[a]	-0.002	0.054	-0.035	0.972
Sex (male)	0.349	0.1	3.487	0.001
Age (juvenile)	-0.92	0.104	-8.887	<0.001

Reduced dataset without animals that are still alive.

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 79.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 248.4 (342.1) sec.

Table S5b: Maze with reduced dataset: results of the linear model testing the effect of cognitive performance (number of errors until criterion) on individuals' survival in days

Predictor variable	Estimate	SE	t	P
Intercept	6.613	0.119	55.752	<0.001
Errors until criterion[a]	0.056	0.056	1	0.322
Sex (male)	0.378	0.11	3.438	0.001
Age (juvenile)	-0.724	0.126	-5.727	<0.001

Reduced dataset without animals that are still alive.

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 61.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 15 (9.2).

Table S6b: Detour-reaching task with reduced dataset: results of the linear model testing the effect of cognitive performance (% correct responses) on individuals' survival in days

Predictor variable	Estimate	SE	t	P
Intercept	6.624	0.141	46.928	<0.001
% correct[a]	-0.045	0.065	-0.692	0.492
Sex (male)	0.384	0.132	2.905	<0.01
Age (juvenile)	-0.761	0.135	-5.627	<0.001

Reduced dataset without animals that are still alive.

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 55.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1; original mean (sd)= 68.4 (22.7) %.

Table S7b: Results of the linear model testing the effects of cognitive performance of all cognitive tests on individuals' survival in days with the reduced dataset

Predictor variable	**Estimate**	**SE**	**t**	**P**
Intercept	6.766	0.191	35.448	<0.001
Food extraction: solving time[a]	-0.253	0.155	-1.633	0.113
String pulling: latency success[a]	0.379	0.177	2.146	0.04
Maze: errors until criterion[a]	-0.011	0.08	-0.138	0.891
Detour reaching: % correct[a]	-0.063	0.081	-0.777	0.443
Sex (male)	0.343	0.171	2.008	0.054
Age (juvenile)	-0.812	0.2	-4.066	<0.001

Reduced dataset without animals that are still alive.

Please note that an outlier individual with extreme high solving time in the food extraction task was excluded from the dataset, as it changed model results significantly.

Reference category for categorical predictor is indicated in brackets, SE: Standard error, N= 37.

[a] Covariate was z-transformed to a mean of= 0 and sd= 1.

General discussion

In this thesis, I focussed on individual variation in different cognitive abilities in a wild primate and linked cognitive performance with important fitness proxies. I investigated the determinants and potential confounds of individual differences in cognitive performance, studied the structure of individuals' performance across different cognitive tasks and linked cognitive variation with different fitness components. Subjects were tested in five cognitive tasks during short-term captivity. The cognitive tasks addressed different ecologically relevant cognitive abilities of grey mouse lemurs. These included associative and motor learning during repeated problem solving, causal reasoning and spontaneous problem solving in a string-pulling task, spatial learning, inhibitory control and behavioural flexibility during reversal learning. In contrast to most other studies that investigated individual variation in a single cognitive performance measure and its correlation with a fitness measure, I subjected the same individuals to several tasks of a cognitive test battery. This allowed to also assess how performance in one cognitive task and domain relates to performance in other domains, thus investigating the structure of cognition. Importantly, as highlighted throughout this thesis, a reliable quantification of individual differences in cognitive abilities is crucial and several pitfalls and challenges have been addressed. In the first part of this general discussion, I will address therefore the most important aspects in this regard, as this provides the foundation for the following discussion of my findings. I will discuss my results on the structure and fitness consequences of cognition in the light of recent empirical findings and theoretical approaches and discuss how my study contributes to the framework for the evolution of cognition. Finally, I will draw general conclusions and suggest some future directions for research on the evolution of cognitive abilities.

Individual cognitive variation as a basis for the evolution of cognition and its study

Similar to other recent studies addressing individual variation in cognitive abilities, mouse lemurs differed widely in the different measures of cognitive performance. Further, the results of this thesis demonstrated, that it is possible to assess personality differences and cognitive performance of relatively large numbers of unhabituated grey mouse lemurs during short-term captivity, a requirement that cannot be taken for granted in a wild animal.

The test performance in the cognitive tasks was not systematically influenced by non-cognitive factors and individual characteristics like age and sex, but also motivation and personality differences (Chapter 1), which is important when quantifying individual cognitive variation (Thornton et al., 2014; Morand-Ferron et al., 2016). In the following, I will discuss the most important aspects in this regard in the light of recent findings and recommendations in more detail.

The choice of cognitive tasks – A problem with problem solving?

The cognitive tasks that were used to assess variation in cognitive abilities in the grey mouse lemurs had to be feasible to be applied to wild subjects during short-term captivity while reaching high sample sizes for the investigation of the adaptive value of cognitive traits. Therefore, typical cognitive tasks that require long phases of training and testing which are commonly applied to captive individuals were not suitable. In order to meet these challenges, I assessed cognitive performance in two problem-solving tasks addressing associative and novel motor learning as well as causal reasoning and spontaneous problem solving. Furthermore, I tested subjects in commonly used cognitive tasks and assessed spatial learning in a maze, inhibitory control in a detour-reaching task and behavioural flexibility during reversal learning.

Problem-solving tasks have been predominantly applied to study cognition in the wild as they do not require training and subjects can voluntarily interact with the task, even in their natural habitat (Keagy et al., 2009; Cole et al., 2012; Cauchard et al., 2013; Preiszner et al., 2017; Wetzel, 2017). However, their usage has been criticized because cognitive processes involved in problem solving are difficult to define (Rowe & Healy, 2014; Thornton et al., 2014; van Horik & Madden, 2016). Further, performance is prone to be influenced by chance or non-cognitive factors like motivation or personality traits (Thornton et al., 2014; for a detailed review of mechanisms please see (Griffin & Guez, 2014; Griffin & Guez, 2016). Thus, problem-solving tasks as means for assessing cognitive performance have to be applied with caution. For example, in pheasants, problem-solving performance in three food extraction tasks did only improve in the most complex task that could not be solved by persistent pecking (i.e.

motivation), but required more precise grasping actions. In the two other less complex problem-solving tasks individuals did not learn across two trials but motivation determined performance (van Horik & Madden, 2016).

To avoid these issues, problem-solving tasks and performance measures in this thesis were designed and chosen to avoid solving by chance or pure persistence: I measured repeated solving efficiency in the food extraction tasks and also success in the string-pulling task required a repeated and directed pulling action. Performance in the two tasks was repeatable and not influenced by systematic non-cognitive factors like body condition, neophilia and activity (Chapter 1). Moreover, repeated solving efficiency in the food extraction task correlated with individuals' learning slopes, suggesting improved motor learning across the six possible trials (Chapter 2, supplementary material). Also, performance in the two problem-solving tasks correlated while motivation to engage with the respective task did not, indicating that similar cognitive processes are involved (Chapter 1). Thus, the results of this thesis demonstrate that using more complex problem-solving designs with carefully chosen performance variables, such as non-binary measures of success and repeated solving opportunities, allows to quantify cognitive performance in wild subjects. Combined with common psychometric tests and the simultaneous study of non-cognitive factors, this provides a good approach to study the cognitive abilities of relatively large numbers of wild individuals. Eventually, using this combination of tasks allows contributing data from new species to the growing body of research that links cognition and fitness in the wild.

Motivation and personality as potentially confounding factors

Motivation influences virtually every behaviour and therefore also performance in a cognitive task, but its mechanistic basis is still not fully understood (Houston & McFarland, 1976; Rowe & Healy, 2014). In order to quantify cognitive performance reliably and allow comparisons across individuals, it is therefore important to assess motivation or its proxies in the context of the respective cognitive task (Rowe & Healy, 2014; Morand-Ferron et al., 2016).

In this thesis, I used body condition (BMI) as a measure of energetic state and proxy for feeding motivation in the food-rewarded tasks. Body condition did not correlate with performance in the string-pulling task, inhibitory control or spatial learning. However, it predicted performance in some measures in the food extraction task but did not correlate with solving efficiency during the repeated food extraction, which was therefore used as the main performance measure in this task (Chapter 1 and 2).

As an additional behavioural measure of motivation, I used task-directed behaviour in the two problem-solving tasks, in which subjects could voluntarily interact with the task apparatuses. This task-directed motivation predicted performance in the food extraction and string-pulling task. However, this is a common finding in motor-learning or problem-solving tasks because only subjects that actively engage with the details of a task are likely to learn or solve them (Griffin & Guez, 2014; and please see discussion in Chapter 1 for details). Interestingly, in the grey mouse lemurs, individuals were not consistent in this task-directed motivation across the two tests and task-directed motivation did not correlate with body condition, personality or other individual characteristics. This could suggest that task-directed motivation itself has a cognitive component. Furthermore, it excludes the possibility that consistent intrinsic motivation accounted for performance across tasks, an important prerequisite when investigating correlations between different cognitive performances in the study of general cognitive abilities (Shaw & Schmelz, 2017).

With the growing interest in individual cognitive variation, also the interest in links between cognition and personality has emerged. Personality traits can covary with the way animals gather and act on information (Carere & Locurto, 2011; Sih & Del Giudice, 2012; Griffin et al., 2015) and could influence learning probabilities. For example, traits like boldness or exploration could influence probabilities or rates of exposure with learning contingencies in nature or task features in a cognitive testing design (Carere & Locurto, 2011; Morand-Ferron et al., 2016; Guillette et al., 2017; example in Damerius et al., 2017). Therefore, assessing personality traits that could cause individual variation in test performance is important to reliably quantify cognitive performance across individuals (Morand-Ferron et al., 2016). In the grey mouse lemurs, the personality traits neophilia (latency to contact a novel object) and activity (duration actively exploring a novel environment during an open field test) did not predict performance in any of the five tasks (Chapter 1). This shows that there was no performance bias due to personality traits in the chosen cognitive performance measures and test designs did not constrain, for example, less neophilic or shy individuals in engaging with and learning in the tasks.

Another aspect in this regard is that especially when testing wild, free-ranging subjects, differences in personality or other individual characteristics may bias the voluntary participation in cognitive experiments (Morand-Ferron et al., 2016; Shaw & Schmelz, 2017). For example, sampling may be biased because shy, less explorative individuals are less likely to participate in freely accessible tasks, or to enter traps for cognitive testing during short-term captivity (Biro & Dingemanse, 2009; Carter et al., 2012; van Horik et al., 2017). Therefore, when subjects are directly tested in their natural habitat, participation rates are usually low (e.g. 47% in meerkats, *Suricata suricatta;* Thornton & Samson, 2012*;* or 6% in great tits, Morand-Ferron et al., 2015). However, in the study population of grey mouse lemurs, capture probabilities and recapture rates are high (Kraus et al., 2008), while drop-out rates of animals, that did not engage with the cognitive tasks and could not be tested, were low. Thus, in contrast to many other studies focusing on few participating individuals often of one sex (e.g. (Isden et al., 2013; Shaw et al., 2015), a sampling bias was unlikely in the individuals that participated in this study.

The structure of cognition: Domain-specific cognitive abilities or general intelligence

Within the cognitive performance measures in this thesis, I did not find support for a general factor explaining variation in cognitive performance across cognitive domains, similar to the general intelligence factor (g) in humans (Chapter 1). Subjects were not consistent in their performance across tasks and performance in the cognitive tasks addressing different cognitive domains did not correlate. In contrast, individuals that performed well in the two problem-solving tasks, i.e. in associative and motor learning as well as causal understanding, tended to score low in the other tasks addressing inhibitory control, spatial learning and behavioural flexibility. As discussed in Chapter 1, individuals' contrasting performance in either problem solving or the other tasks might indicate a trade-off in foraging strategies. Individuals could compensate inferior abilities in spatial learning, inhibitory control and behavioural flexibility during foraging with better extractive and innovative foraging. The finding that efficient solving during repeated problem solving, i.e. motor and associative learning in the food extraction task, correlated with the ability to maintain body mass during the harsh dry season supports its role in the foraging context (Chapter 2). In a similar manner, inhibitory control and motor-related abilities were suggested to be involved in a trade-off during foraging in pheasants, where performance in the respective tasks correlated negatively and no general factor explaining cognitive performance across nine tasks could be extracted (van Horik et al., 2018). Also in Carib grackles (*Quiscalus lugubris fortirostris*), speed in problem solving correlated negatively with discrimination and reversal learning abilities and individuals' better performance in only one of both was suggested to reflect different cognitive strategies (Ducatez et al., 2019).

In humans, evidence for domain-general intelligence and the presence of a general intelligence factor g that summarizes performance across cognitive tests is very strong (Deary et al., 2010). Unless general intelligence is inseparably linked to language, it should have evolved in other nonhuman animals as well, especially in closely related species like primates (Burkart et al., 2017). While domain-specific cognitive abilities are thought to have evolved independently in response to specific socio-ecological challenges, domain-general processes, i.e. general intelligence, allow to solve problems flexibly across domains and contexts and could coexist (Amici et al., 2017; Burkart et al., 2017). However, evidence for general intelligence in nonhuman animals is mixed; there is evidence for and against a general factor (g) explaining variation in cognitive performance in all tested taxa or even species so far (please see Chapter 1 for more details).

Especially for primates, evidence for g is contradictory, which make it currently difficult to draw general conclusions on how and when general intelligence evolved. In comparative approaches, different species are compared in their cognitive abilities across cognitive domains. These studies investigate whether some species evolved specialized cognitive skills or rather domain-general intelligence and thus outperform other species across different cognitive tasks (Burkart et al., 2017). Evidence for such general intelligence on the interspecific level (G) was provided in several studies on different primate species, demonstrating that species that performed well in one domain also performed well in others (24 species: Deaner et al., 2006; 62 species: Reader et al., 2011). Mixed approaches, however, in which test batteries are applied to multiple individuals from several species and combined for analysis, did not provide evidence for domain-general intelligence (seven primate species: Amici et al., 2012; four great ape species: Herrmann & Call, 2012). In contrast, Amici et al. (2012) concluded that domain-specific cognitive skills underlie different evolutionary pressures in the different species that led to an (at least partially) modular primate mind.

Finally, a handful of studies on the individual level tested the performance of relatively large amounts of captive individuals, for example with the primate cognition test battery that consists of a broad range of tasks in the physical and social domain (Herrmann et al., 2007). In great apes, evidence for g was supported for orangutans (Damerius et al., 2018) and in one study of chimpanzees (Hopkins et al., 2014). In contrast, cognitive performance was structured in different factors in another study on chimpanzees and human children (Herrmann et al., 2010), thus not providing evidence for g. In New World monkeys, the structure of cognition was assessed in cotton-top tamarins and evidence for a general factor explaining variation in cognitive performance of 22 individuals across 11 tasks was reported (Banerjee et al., 2009). Finally, in a lemur species, which represent the most basal living primates (Yoder, 2007), my study could not find evidence for g in wild grey mouse lemurs.

An important issue in the study of general intelligence in animals is that, in contrast to humans, the composition of test batteries varies significantly between study species and cognitive abilities have to be inferred from performance in the respective tasks (Huber, 2017). Therefore, the mixed evidence for g might be partly explained by several inconsistencies across studies that could influence the probability of detecting g (Burkart et al., 2017; van Horik et al., 2018), which I will address in the following. Studies differ largely in the number and kinds of tests, the assessed cognitive abilities and domains, but also whether they control for non-cognitive factors that could systematically affect cognitive performance across tasks (Burkart et al., 2017; Shaw & Schmelz, 2017). For example, if all tests in a test battery involve associative learning albeit in different contexts, correlations between task performances will be more likely and thus also the likelihood to detect a general factor of cognitive performance increases (Herrmann & Call, 2012; van Horik & Lea, 2017). Moreover, statistical methods used in the detection of g vary and most animal studies are limited in sample size which results in low statistical power. For instance, particularly with wild subjects, sample sizes of about 20 individuals are common. Hence, to improve these issues in the future, the challenge will be to design broad cognitive test batteries that address various cognitive domains and can test relatively large sample sizes. Ideally, these test batteries will be applicable to different species and also confounding factors of cognitive performance are controlled for (Shaw & Schmelz, 2017). Furthermore, testing cognitive abilities in the same cognitive domain with several different tests will help to validate cognitive tasks and ensure that abilities in putative cognitive domains are really assessed (Shaw & Schmelz, 2017; Boogert et al., 2018; Völter et al., 2018).

In the analysis on the structure of cognitive performance in wild grey mouse lemurs, I controlled for the effect of potential confounding factors like motivation and personality and followed recent recommendations concerning the statistical analysis (Burkart et al., 2017). However, the number of different tasks and cognitive abilities that I could assess in this thesis as well as the number of individuals that participated in all tasks was limited. Thus, my findings are a valuable starting point to investigate potential general intelligence in grey mouse lemurs. Further data on intra-individual differences in cognitive abilities spanning more cognitive domains, but also tapping into the same domains will help to characterize the structure of cognitive abilities in grey mouse lemurs more fully.

In conclusion, evidence for a g factor in non-human animals is mixed but the positive findings in rodents and primates suggest that domain-general cognitive abilities may not only be present in humans (Burkart et al., 2017). In the future, more carefully designed studies in various species will be needed to validate these findings and to better understand why, how and when general intelligence evolved.

Fitness correlates of cognitive abilities – Implications for the evolution of cognition

Understanding the fitness consequences of individual variation in cognitive abilities is one of the most pressing goals in order to shed light on the evolution of cognition. If we can link cognitive performance with fitness outcomes, this provides a starting point to understand the complex trade-offs, costs and benefits of cognitive traits, that are shaping individual differences in cognitive abilities and lead to their evolution.

In chapter 2 and 3, I linked cognitive performance in the four main tasks with two different fitness proxies: body condition (BMI) change across the harsh dry season and survival or longevity. I found that the ability to efficiently solve the repeated food extraction task correlated with BMI change; individuals that were quick in associating a successful motor action with a reward and quickly repeated this action were better in maintaining their body condition during the dry season in the wild. Interestingly, this was not the case for spatial learning ability (Chapter 2), string-pulling ability and inhibitory control in the detour-reaching task (please see Appendix). For survival and longevity, I did not detect a significant correlation with cognitive performance independent of assessed cognitive ability. In order to expand the diversity of fitness proxies, I investigated individuals’ hair cortisol concentration and hematocrit levels which could be summarized together with body condition into a physiological condition factor. However, this condition factor did not predict short-term survival of subjects and was therefore not used as a fitness proxy in this thesis (please see discussion chapter 3 for potential reasons). Further investigations into the physiological condition indices and their link with survival and other fitness proxies will be needed to draw final conclusions on their utility as a fitness proxy. Nevertheless, the difficulty in validating the measure as a fitness proxy further demonstrates that identifying and operationalizing meaningful fitness, that can be studied together with cognitive abilities in the wild, is challenging for most taxa. Together with the challenges of cognitive testing, this might explain why the entire class of mammals remains largely underrepresented with only two studies on fitness correlates of cognitive abilities in the wild (this study and Maille & Schradin, 2016 on African striped mice).

Foraging success as a mediator between cognitive abilities and fitness outcomes

A common mechanism by which cognitive processes are hypothesized to improve fitness is by increasing foraging efficiency or quality which impacts body condition and growth (e.g. Dukas & Bernays, 2000; Roth et al., 2010). However, assessing measures of body condition directly but also observing foraging success is difficult for most taxa in the wild, which is why different proxies have been used. Colony foraging success correlated positively with learning speed in bumble bee colonies (Raine & Chittka, 2008) but not on the individual bee level (Evans et al., 2017). For birds, foraging success is meaningful in regard to the individual's own body condition but it also directly affects the feeding of nestlings and is therefore likely to predict reproductive success (Ydenberg, 1994). In great tits, problem-solving success correlated positively with reproductive success (Cole et al., 2012; Cauchard et al., 2013). Moreover, successful problem solvers had a shorter daily timespan of provisioning the young and smaller home ranges, while provisioning rate was higher compared to birds that did not solve the novel problem. This suggests that successful problem solvers could increase their reproductive success because they were more efficient at exploiting their environment (Cole et al., 2012; Cauchard et al., 2017). By contrast, general cognitive performance also correlated with reproductive success in Australian magpie females, but there was no relationship between cognitive performance and foraging efficiency (Ashton et al., 2018).

In grey mouse lemurs, observing nocturnal foraging is difficult but assessing body measures of focal subjects is feasible. The ability to maintain body condition during the long dry season is especially meaningful, as food is scare (Schmid, 1999; Dammhahn & Kappeler, 2008a). Moreover, at the end of the dry season, the mating season takes place during which body condition influences males’ mating success (Eberle & Kappeler, 2004b) and females’ mating strategies (Huchard et al., 2012). Therefore, I used body condition (BMI) change across the dry season as a fitness proxy: individuals that did not change much in BMI in this time either used less or were able to acquire more energy compared to individuals whose BMI decreased in the same period. Performance in spatial learning, causal reasoning during spontaneous string pulling and inhibitory did not predict BMI change across the dry season. This suggests that these cognitive abilities did not help mouse lemurs to maintain their body condition during the dry season by saving energy or improved foraging. Efficient repeated solving in the food extraction task predicted the ability to maintain body condition across the harsh dry season, however. This positive correlation indicates that innovative foraging might be also ecologically meaningful during food scarce seasons in the wild and that involved cognitive abilities could determine an important predictor of fitness in grey mouse lemurs.

Why positive correlations between cognitive abilities and measures of fitness might not always be expected

Measures of survival are an important predictor of fitness as survival until breeding and across multiple breeding seasons determines the chance to reproduce (Blums et al., 2005; Wilson & Nussey, 2010). However, in the study of fitness consequences of cognitive traits, so far only survival in the short-term has been used as a fitness proxy. Positive as well as negative correlations between cognitive performance and survival in the same species have been found (dependent on individuals' sex: Maille & Schradin, 2016; dependent on the cognitive measure and body mass of individuals: Madden et al., 2018), as well as no significant link (Cole et al., 2012) has been detected. Thus, these contrasting findings already indicate, that links between cognitive abilities and survival are complex and likely to depend on various variables.

In the grey mouse lemurs, I did not detect a link between cognitive performance, regardless of the addressed cognitive ability, and survival or longevity. As reviewed in the introduction, this absence of a link fits into the overall complex, ambiguous picture of correlations between cognitive performance and fitness proxies across studies and species. Yet, the difficulty in detecting a link between cognitive abilities and fitness proxies or possible absence of a link can have several reasons, which I will summarize in the following and discuss in the light of recent empirical findings and theoretical frameworks.

First of all, addressing methodological concerns, cognitive abilities have to be reliably quantified, a topic that I addressed in detail in Chapter 1 and at the beginning of this general discussion and could confirm within the assessed non-cognitive factors and repeatability measures. Moreover, a lack of statistical power could lead to non-significant findings as large sample sizes are needed in selection analyses (Kingsolver et al., 2001; Dingemanse & Reale, 2005). However, the sample size in this thesis with 40 – 86 individuals, depending on the cognitive measure, was comparable to other studies in which cognitive performance and a measure of fitness correlated.

Excluding these methodological issues, it is possible that tested cognitive abilities do not provide individuals with fitness benefits or benefits in the assessed fitness proxy are masked by other uncorrelated factors (Morand-Ferron et al., 2016). The assessed cognitive abilities in my thesis are likely to impact survival via a better maintenance of body condition, e.g. due to better memory of locations of food resources or better extractive foraging abilities, As mentioned before, I could support the role of extractive foraging abilities in maintaining body condition during the dry season, but nevertheless this did not seem to translate into better survival. However, grey mouse lemur’ survival is also crucially impacted by predation (Goodman et al., 1993; Fichtel, 2016; and details in Chapter 2 and 3) which could mask a potentially positive effect of the assessed cognitive abilities on survival. Thus, if the addressed cognitive abilities do not provide benefits in avoiding, detecting and escaping predators, this could explain the difficulty to detect links with longevity. Assessing cognitive abilities involved in predator detection and avoidance would, therefore, be a promising yet challenging study in the future.

Alternatively, fitness benefits of cognitive abilities might be counteracted by direct fitness costs of better cognitive capacities or by other correlated traits that reduce fitness (Morand-Ferron et al., 2016). First of all, neuronal tissue and information gathering and processing itself is energetically costly (Laughlin et al., 1998; Kawecki, 2010; Kotrschal et al., 2013). In artificial selection experiments, increased learning ability correlated with a reduction in immunity in laboratory mice (*Mus musculus*) (Barnard et al., 2006) and reduced longevity

(Burger et al., 2008) and larval competitive ability (Mery & Kawecki, 2003) in fruit flies (*Drosophila melanogaster*). For natural populations, these fitness trade-offs are much more difficult to detect. Yet, a study in butterflies (*Pieris rapae*) found that better learning correlated with fewer and less developed eggs (Snell-Rood et al., 2011). Moreover, cognitive traits might be correlated with other behaviours that have fitness disadvantages and therefore increased cognitive abilities do not result in net benefit of fitness. For example, individual variation in cognitive ability was found to be linked to stress reactivity, boldness and exploration (reviewed in Øverli et al., 2007; Carere & Locurto, 2011). In wild great tits, individuals with higher cognitive performance were less able to compete for food (Cole & Quinn, 2012), and more likely to desert their nests with dependent offspring, probably due to higher sensitivity to perceived predation risk during experimental trapping and handling (Cole et al., 2012). Unfortunately, detecting these correlated costs is extremely challenging in the wild, especially in a small nocturnal primate, for which behavioural observations are difficult. I could test links with personality, however, but did not find that cognitive abilities correlated with the assessed personality traits (Chapter 1). Thus, it seems unlikely that individuals that performed better in the cognitive tasks were, for example, more neophilic or bold which could have reduced sensitivity to predators and negatively affected survival (e.g. Hulthén et al., 2017).

Cognitive performance in the grey mouse lemurs could not be summarized into one general factor but individuals differed in performance across tasks (Chapter 1). Therefore, trade-offs between cognitive abilities within individuals could result into equally effective cognitive strategies which would explain the missing correlation between cognitive performance in a given task and longevity (Rowe & Healy, 2014; Morand-Ferron et al., 2016). Only two other studies assessed performance of more than two different cognitive abilities and linked them with fitness outcomes. However, these studies did not detect different cognitive strategies but performance across tasks correlated and could be summarized into a general factor. This factor predicted reproductive success in Australian magpies (Ashton et al., 2018) but did not correlate with mating success in bower birds (Isden et al., 2013). Hence, more studies that assess intra-individual differences in cognitive abilities and link them with fitness outcomes in the wild will be needed to better understand these potential cognitive strategies or trade-offs and their fitness consequences in the future.

Finally, what is beneficial and a “better” cognitive ability depends on the time, space, i.e. environmental and social context (Rowe & Healy, 2014; Ten Cate, 2014; Morand-Ferron et al., 2016). Together with the complex relationships between various traits and involved fitness trade-offs this may explain why individual variation in cognitive traits is maintained and selection of cognitive traits is so difficult to detect in natural populations (Rowe & Healy, 2014; Morand-Ferron et al., 2016; Boogert et al., 2018). Ultimately, also in humans, the advantages of better general cognitive abilities are not as overarching as perhaps expected. In our modern human performance societies, intelligence predicts indeed various important life outcomes, such as socioeconomic success, health and survival, but was also found to correlate negatively with reproductive success, the most direct predictor of evolutionary fitness (Strenze, 2006; Shatz, 2008; Reeve et al., 2013; Plomin & Deary, 2015). Thus, this thesis, together with the other studies until today, reveals that cognition and fitness are not necessarily positively correlated, but that involved relationships are complex and we are still in the beginning to understand how cognition evolved.

Conclusions and future directions

My thesis demonstrated that it is feasible to study cognitive abilities in different cognitive tasks with relatively large sample sizes in a wild primate and link performance with different fitness outcomes. I could show that individuals varied substantially in their cognitive performance and that performance was not systematically affected by non-cognitive factors which is important to exclude, especially, but not exclusively when testing wild subjects. In respect of the lately criticized use of problem-solving tasks, my results demonstrated that when carefully designing these tasks and thoughtfully choosing performance measures, they provide a useful tool to assess repeatable cognitive performance of relatively large numbers of wild subjects. I investigated intra-individual variation in cognitive performance across the different tasks and abilities and did not find evidence for a general factor explaining cognitive performance in grey mouse lemurs. Thus, my study contributes valuable findings for the ongoing question when, how and why general cognitive abilities evolved.

Most importantly, this thesis contributed the first study in a wild primate to the growing body of research on the adaptive value of cognitive traits. My results showed that better cognitive performance during efficient problem solving correlated with an important fitness proxy in the short-term, i.e. the maintenance of body condition during food scare conditions. However, better cognitive abilities did not translate into better long-term fitness benefits, such as survival and longevity. Thus, together with the present body of research revealing heterogeneous links between cognition and fitness across cognitive and fitness measures as well as species, my findings demonstrated that cognitive abilities are involved in complex interactions between various traits. At the same time, they are likely to have cost and benefits and thus do not necessarily result in positive correlations with fitness outcomes. Unravelling this complex system, the evolution of cognition, will require more studies on individual differences in cognitive abilities and their link with fitness in the wild. Some suggestions for these future studies, I will discuss in the following.

Concerning the link between cognition and fitness in grey mouse lemurs, it will be particularly interesting to link cognitive performance with measures of reproductive success. This would also allow to test whether the correlation between cognitive performance and maintenance of body condition in the food-scarce season translates into reproductive success. As discussed before, the ability to maintain body condition during the dry season is likely to influence the body mass dependent mating success of males and reproductive strategies of females during the subsequent mating season. In addition, it is promising to design tasks that address cognitive abilities involved in predator detection and avoidance in order to test the link between performance in these tasks and survival. Also, further investigation of the structure of cognitive abilities in grey mouse lemurs (either in the laboratory or wild) will be important to get more evidence for the absence or presences of a general intelligence factor in the species and a better understanding of its evolution within primates. In this regard, it will be promising to increase the number of cognitive domains but also to further validate the assessed cognitive domains by testing the same subjects in different cognitive tasks addressing the same cognitive domain.

To further expand our knowledge on the adaptive value of cognitive traits in the wild, it will be especially promising to expand the number of different cognitive abilities, potential confounding covariates, as well as fitness proxies that are measured within a given species. This will allow a more complete picture of trait interactions, complex relationships and involved trade-offs. My study also revealed, that identifying and operationalizing meaningful proxies can be challenging for some taxa. Therefore, further detailed studies of fitness proxies in different species will help to expand the body of research so far, that mainly focused on different species of birds. The need to establish broad cognitive test batteries, ideally applicable to numerous individuals and a variety of species in the wild, has been highlighted before (Thornton et al., 2014; Shaw & Schmelz, 2017). However, its implementation still proves to be challenging. Cooperation across field sites and automatic testing devices, that allow to test voluntarily participating, individually identifiable subjects in their natural habitat, will be certainly promising in this regard (Morand-Ferron et al., 2016). So far, no study has linked cognitive abilities of the social domain, such as social learning, with fitness outcomes. This field of research could provide crucial insight into the social intelligence hypothesis (Byrne & Whiten, 1988) but is certainly associated with challenges in designing feasibly cognitive tasks. Once these ambiguous goals are achieved and the missing pieces of the puzzle of inter- and intra-individual differences in cognitive abilities and their fitness consequences are gained, a more complete understanding on the evolution of cognition can begin to emerge.

Appendix

Relationship between performance in the string-pulling and detour-reaching task and BMI change across the dry season

Subjects' latency until success in the string-pulling task did not predict their BMI change across the dry season (full null model comparisons: $F_{1,40}$= 0.035, P= 0.852, Table A1). Also subjects' performance in the detour-reaching task, their percentage of correct trials did not significantly predict their BMI change (full null model comparisons: $F_{1,28}$= 1.83, P= 0.190, Table A1).

Table A1: Results of the linear models (LM) fitting the effects of test performance in string-pulling task and detour-reaching task on BMI change from the rainy to the end of dry season

	Predictor variable	**Estimate**	**SE**	**t**	**P**
Model 1: String-pulling task N= 44	Intercept	0.46	0.07	6.26	<0.001
	Latency success[a]	-0.01	0.05	-0.19	0.852
	Sex (male)	-0.41	0.10	-4.19	<0.001
	Age (juvenile)	-0.06	0.11	-0.59	0.559
Model 2: Detour-reaching task N= 32	Intercept	0.67	0.14	4.83	<0.001
	% correct[b]	-0.11	0.08	-1.36	0.186
	Sex (male)	-0.52	0.15	-3.44	0.002
	Age (juvenile)	-0.25	0.16	-1.61	0.118

The statistical analysis was conducted in the same manner as the analysis of the effects of cognitive performance on BMI change in Chapter 2, please see methods there.

Reference categories for categorical predictors are indicated in brackets, SE: Standard error.

[a] Covariate was log transformed and afterwards z-transformed to a mean of= 0 and sd= 1

[b] Covariate was z-transformed to a mean of= 0 and sd=1

References

Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Current Biology*, *18*(18), 1415–1419.

Amici, F., Barney, B., Johnson, V. E., Call, J., & Aureli, F. (2012). A modular mind? A test using individual data from seven primate species. *PLoS ONE*, *7*(12), e51918.

Amici, F., Call, J., & Aureli, F. (2017). Coexistence of general intelligence and specialized modules. *Behavioral and Brain Sciences*, *40*, e196.

Arden, R., & Adams, M. J. (2016). A general intelligence factor in dogs. *Intelligence*, *55*, 79–85.

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.

Balda, R. P., & Kamil, A. C. (1989). A comparative study of cache recovery by three corvid species. *Animal Behaviour*, *38*(3), 486–495.

Banerjee, K., Chabris, C. F., Johnson, V. E., Lee, J. J., Tsao, F., & Hauser, M. D. (2009). General intelligence in another primate: Individual differences across cognitive task performance in a new world monkey (Saguinus oedipus). *PLoS ONE*, *4*(6).

Barkley, C. L., & Jacobs, L. F. (2007). Sex and species differences in spatial memory in food-storing kangaroo rats. *Animal Behaviour*, *73*(2), 321–329.

Barnard, C. J., Collins, S. A., Daisley, J. N., & Behnke, J. M. (2006). Odour learning and immunity costs in mice. *Behavioural Processes*, *72*(1), 74–83.

Barnett, C. A., Suzuki, T. N., Sakaluk, S. K., & Thompson, C. F. (2015). Mass-based condition measures and their relationship with fitness: in what condition is condition? *Journal of Zoology*, *296*(1), 1–5.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3).

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Beehner, J. C., & Bergman, T. J. (2017). The next step for stress research in primates: To identify relationships between glucocorticoid secretion and fitness. *Hormones and Behavior*, *91*, 68–83.

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.

Billett, H. H. (1990). Hemoglobin and hematocrit. In H. Walker, W. Hall, & J. Hurst (Eds.), *Clinical methods: The history, physical, and laboratory examinations.* Boston: Butterworths.

Birchard, G. F. (1997). Optimal hematocrit: Theory, regulation and implications. *American Zoologist*, *37*(1), 65–72.

Biro, P. A., & Dingemanse, N. J. (2009). Sampling bias resulting from animal personality. *Trends in Ecology & Evolution*, *24*(2), 66–67.

Blums, P., Nichols, J. D., Hines, J. E., Lindberg, M. S., & Mednis, A. (2005). Individual quality, survival variation and patterns of phenotypic selection on body condition and timing of nesting in birds. *Oecologia*, *143*(3), 365–376.

Bond, A. B., Kamil, A. C., & Balda, R. P. (2007). Serial reversal learning and the evolution of behavioral flexibility in three species of North American corvids (Gymnorhinus cyanocephalus, Nucifraga columbiana, Aphelocoma californica). *Journal of Comparative Psychology*, *121*(4), 372–379.

Bonier, F., Martin, P. R., Moore, I. T., & Wingfield, J. C. (2009). Do baseline glucocorticoids predict fitness? *Trends in Ecology & Evolution*, *24*(11), 634–642.

Bonier, F., Moore, I. T., Martin, P. R., & Robertson, R. J. (2009). The relationship between fitness and baseline glucocorticoids in a passerine bird. *General and Comparative Endocrinology*, *163*(1–2), 208–213.

Boogert, N. J., Anderson, R. C., Peters, S., Searcy, W. A., & Nowicki, S. (2011). Song repertoire size in male song sparrows correlates with detour reaching, but not with other cognitive measures. *Animal Behaviour*, *81*(6), 1209–1216.

Boogert, N. J., Fawcett, T. W., & Lefebvre, L. (2011). Mate choice for cognitive traits: A review of the evidence in nonhuman vertebrates. *Behavioral Ecology*, *22*(3), 447–459.

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).

Boogert, N. J., Monceau, K., & Lefebvre, L. (2010). A field test of behavioural flexibility in Zenaida doves (Zenaida aurita). *Behavioural Processes*, *85*(2), 135–141.

Bowers, E. K., Hodges, C. J., Forsman, A. M., Vogel, L. A., Masters, B. S., Johnson, B. G. P., … Sakaluk, S. K. (2014). Neonatal body condition, immune responsiveness, and hematocrit predict longevity in a wild bird population. *Ecology*, *95*(11), 3027–3034.

Brown, C., & Braithwaite, V. A. (2005). Effects of predation pressure on the cognitive ability of the poeciliid Brachyraphis episcopi. *Behavioral Ecology*, *16*(2), 482–487.

Bshary, R., Wickler, W., & Fricke, H. (2002). Fish cognition: a primate's eye view. *Animal Cognition*, *5*(1), 1–13.

Budaev, S. V. (2010). Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, *116*(5), 472–480.

Burger, J. M. S., Kolss, M., Pont, J., & Kawecki, T. J. (2008). Learning ability and longevity: A symmetrical evolutionary trade-off in drosophila. *Evolution*, *62*(6), 1294–1304.

Burkart, J. M., Schubiger, M. N., & van Schaik, C. P. (2017). The evolution of general intelligence. *Behavioral and Brain Sciences*, *40*, e195.

Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence : social expertise and the evolution of intellect in monkeys, apes, and humans*. Clarendon Press.

Call, J., & Tomasello, M. (1996). The effect of humans on the cognitive development of apes. In A. E. Russon, S. T. Parker, & K. A. Bard (Eds.), *Reaching into Thought: The Minds of the Great Apes.* (pp. 371–403). New York, NY, US: Cambridge University Press.

Campbell, T. W. (1994). Cytology. In B. W. Ritchie, G. J. Harrison, & L. R. Harrison (Eds.), *Avian Medicine: Principles and Applications* (pp. 199–221). Lake Worth: Winger.

Carere, C., & Locurto, C. (2011). Interaction between animal personality and animal cognition. *Current Zoology*, *57*(4), 491–498.

Carter, A. J., Heinsohn, R., Goldizen, A. W., & Biro, P. A. (2012). Boldness, trappability and sampling bias in wild lizards. *Animal Behaviour*, *83*(4), 1051–1058.

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*, 1–8.

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.

Cauchoix, M., & Chaine, A. S. (2016). How can we study the evolution of animal minds?

Frontiers in Psychology, *7*, 1–18.

Cauchoix, M., Chow, P. K. Y., Horik, J. O. van, 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), 20170281.

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*, 273–283.

Clayton, N. S. (2004). Is necessity the mother of innovation? *Trends in Cognitive Sciences*, *8*(3), 98–99.

Clutton-Brock, T. H., & Harvey, P. H. (1980). Primates, brains and ecology. *Journal of Zoology*, *190*(3), 309–323.

Clutton-Brock, T. H., Major, M., Albon, S. D., & Guinness, F. E. (1987). Early development and population dynamics in red deer. I. Density-dependent effects on juvenile survival. *The Journal of Animal Ecology*, *56*(1), 53.

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.

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.

Coop, R. L., & Kyriazakis, I. (1999). Nutrition-parasite interaction. *Veterinary Parasitology*, *84*(3–4), 187–204.

Coppens, C. M., de Boer, S. F., & Koolhaas, J. M. (2010). Coping styles and behavioural flexibility: towards underlying mechanisms. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1560), 4021–4028.

Croston, R., Branch, C. L., Kozlovsky, D. Y., Dukas, R., & Pravosudov, V. V. (2015). Heritability and the evolution of cognitive traits. *Behavioral Ecology*, *26*(6), 1447–1459.

D’Hooge, R., & De Deyn, P. P. (2001). Applications of the Morris water maze in the study of learning and memory. *Brain Research. Brain Research Reviews*, *36*(1), 60–90.

Dall, S. R. X., & Griffith, S. C. (2014). An empiricist guide to animal personality variation in ecology and evolution. *Frontiers in Ecology and Evolution*, *2*, 1–7.

Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004). The behavioural ecology of personality: Consistent individual differences from an adaptive perspective. *Ecology Letters*, *7*(8), 734–739.

Damerius, L. A., Burkart, J. M., van Noordwijk, M. A., Haun, D. B. M., Kosonen, Z. K., Galdikas, B. M. F., … van Schaik, C. P. (2018). General cognitive abilities in orangutans (Pongo abelii and Pongo pygmaeus). *Intelligence*.

Damerius, L. A., Graber, S. M., Willems, E. P., & van Schaik, C. P. (2017). Curiosity boosts orang-utan problem-solving ability. *Animal Behaviour*, *134*, 57–70.

Dammhahn, M. (2012). Are personality differences in a small iteroparous mammal maintained by a life-history trade-off? *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2645–2651.

Dammhahn, M., & Almeling, L. (2012). Is risk taking during foraging a personality trait? A field test for cross-context consistency in boldness. *Animal Behaviour*, *84*(5), 131–1139.

Dammhahn, M., & Kappeler, P. M. (2008a). Comparative feeding ecology of sympatric Microcebus berthae and M. murinus. *International Journal of Primatology*, *29*(6), 1567–1589.

Dammhahn, M., & Kappeler, P. M. (2008b). Small-scale coexistence of two mouse lemur species (Microcebus berthae and M. murinus) within a homogeneous competitive environment. *Oecologia*, *157*(3), 473–483.

Darwin, C. (1859). On the origin of the species by means of natural selection. In *On the Origin of the Species by Means of Natural Selection*. London, UK: John Murray.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London,: John Murray.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., … Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*(10), 996–1005.

Deaner, R. O., Schaik, C. P. Van, & Johnson, V. (2006). Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evolutionary Psychology*, *4*, 149–196.

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*(3), 201–211.

Dechmann, D. K. N., & Safi, K. (2009). Comparative studies of brain evolution: a critical insight from the Chiroptera. *Biological Reviews of the Cambridge Philosophical Society*, *84*(1), 161–172.

Dingemanse, N. J., & Reale, D. (2005). Natural selection and animal personality. *Behaviour*, *142*(9), 1159–1184.

Domjan, M., Cusato, B., & Villarreal, R. (2000). Pavlovian feed-forward mechanisms in the control of social behavior. *The Behavioral and Brain Sciences*, *23*(2), 235-49; discussion 249-82.

Dougherty, L. R., & Guillette, L. M. (2018). Linking personality and cognition: a meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756), 20170282.

Ducatez, S., Audet, J.-N., & Lefebvre, L. (2019). Speed–accuracy trade-off, detour reaching and response to PHA in Carib grackles. *Animal Cognition*, 1–9.

Dukas, R. (2004). Evolutionary biology of animal cognition. *Annual Review of Ecology, Evolution, and Systematics*, *35*(1), 347–374.

Dukas, R. (2008). Evolutionary biology of insect learning. *Annual Review of Entomology*, *53*(1), 145–160.

Dukas, R., & Bernays, E. A. (2000). Learning improves growth rate in grasshoppers. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(6), 2637–2640.

Dukas, R., & Visscher, P. K. (1994). Lifetime learning by foraging honey bees. *Animal Behaviour*, *48*(5), 1007–1012.

Dunbar, R. I. ., & Shultz, S. (2007). Understanding primate brain evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 649–658.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190.

Dyer, F. (1998). Cognitive ecology of navigation. In R. Dukas (Ed.), *Cognitive Ecology.* (pp. 201–260). Chicago: University of Chicago Press.

Eberle, M., & Kappeler, P. (2002). Mouse lemurs in space and time: a test of the socioecological model. *Behavioral Ecology and Sociobiology*, *51*(2), 131–139.

Eberle, M., & Kappeler, P. M. (2004a). Selected polyandry: Female choice and inter-sexual conflict in a small nocturnal solitary primate (Microcebus murinus). *Behavioral Ecology and Sociobiology*, *57*(1), 91–100.

Eberle, M., & Kappeler, P. M. (2004b). Sex in the dark: Determinants and consequences of mixed male mating tactics in Microcebus murinus, a small solitary nocturnal primate.

Behavioral Ecology and Sociobiology, *57*(1), 77–90.

Eberle, M., & Kappeler, P. M. (2006). Family insurance: kin selection and cooperative breeding in a solitary primate (Microcebus murinus). *Behavioral Ecology and Sociobiology*, *60*(4), 582–588.

Ellegren, H., & Sheldon, B. C. (2008). Genetic basis of fitness differences in natural populations. *Nature*, *452*(7184), 169–175.

Emery, N. J., & Clayton, N. S. (2004). The Mentality of Crows: Convergent Evolution of Intelligence in Corvids and Apes. *Science*, *306*(5703), 1903–1907.

Ethan Pride, R. (2005). High faecal glucocorticoid levels predict mortality in ring-tailed lemurs (Lemur catta). *Biology Letters*, *1*(1), 60–63.

Evans, L. J., Smith, K. E., & Raine, N. E. (2017). Fast learning in free-foraging bumble bees is negatively correlated with lifetime resource collection. *Scientific Reports*, *7*(1), 1–10.

Fair, J., Whitaker, S., & Pearson, B. (2007). Sources of variation in haematocrit in birds. *Ibis*, *149*(3), 535–552.

Farand, É., Allainé, D., & Coulon, J. (2002). Variation in survival rates for the alpine marmot (*Marmota marmota*): effects of sex, age, year, and climatic factors. *Canadian Journal of Zoology*, *80*(2), 342–349.

Fichtel, C. (2012). Predation. In J. Mitani, J. Call, P. Kappeler, R. Palombit, & J. Silk (Eds.), *The evolution of primate societies* (pp. 169–194). Chicago: Chicago University Press.

Fichtel, C. (2016). Predation in the dark: antipredator strategies of Cheirogaleidae and other nocturnal primates. In S. M. Lehman, U. Radespiel, & E. Zimmermann (Eds.), *The Dwarf and Mouse Lemurs of Madagascar* (pp. 366–380). Cambridge: Cambridge University Press.

Fourie, N. H., Brown, J. L., Jolly, C. J., Phillips-Conroy, J. E., Rogers, J., & Bernstein, R. M. (2016). Sources of variation in hair cortisol in wild and captive non-human primates. *Zoology*, *119*(2), 119–125.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*(417), 178.

Fox, J., & Weisberg, S. (2011). An {R} companion to applied regression, 2nd edn. Thousand Oaks, CA: Sage. R package version 2.1.4.

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.

Galsworthy, M. J., Paya-Cano, J. L., Liu, L., Monleón, S., Gregoryan, G., Fernandes, C., … Plomin, R. (2005). Assessing reliability, heritability and general cognitive ability in a battery of cognitive tasks for laboratory mice. *Behavior Genetics*, *35*(5), 675–692.

Gao, W., Stalder, T., Foley, P., Rauh, M., Deng, H., & Kirschbaum, C. (2013). Quantitative analysis of steroid hormones in human hair using a column-switching LC–APCI–MS/MS assay. *Journal of Chromatography B*, *928*, 1–8.

Gonzalez-Voyer, A., Winberg, S., & Kolm, N. (2009). Social fishes and single mothers: brain evolution in African cichlids. *Proceedings. Biological Sciences*, *276*(1654), 161–167.

Goodman, S. M., O'Connor, S., & Langrand, O. (1993). A review of predation on lemurs: Implications for the evolution of social behavior in small, nocturnal primates. In P. Kappeler & J. Ganzhorn (Eds.), *Lemur social systems and their ecological basis* (pp. 51–66). Boston, MA: Springer US.

Greenberg, R. (2003). The role of neophobia and neophilia in the development of innovative behaviour of birds. In S. Reader & K. Laland (Eds.), *Animal innovation* (pp. 175–196). New York: Oxford University Press.

Griffin, A. S. (2004). Social learning about predators: a review and prospectus. *Learning & Behavior*, *32*(1), 131–140.

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 B: Biological Sciences*, *371*(1690).

Griffin, A. S., Diquelou, M., & Perea, M. (2014). Innovative problem solving in birds: A key role of motor diversity. *Animal Behaviour*, *92*, 221–227.

Griffin, A. S., & Guez, D. (2014). Innovation and problem solving: A review of common mechanisms. *Behavioural Processes*, *109*, 121–134.

Griffin, A. S., & Guez, D. (2016). Bridging the gap between cross-taxon and within-species analyses of behavioral innovations in birds: Making sense of discrepant cognition–innovation relationships and the role of motor diversity. *Advances in the Study of Behavior, 48*, 1–40.

Griffin, A. S., Guillette, L. M., & Healy, S. D. (2015). Cognition and personality: An analysis of an emerging field. *Trends in Ecology and Evolution*, *30*(4), 207–214.

Guillette, L. M., Naguib, M., & Griffin, A. S. (2017). Individual differences in cognition and personality. *Behavioural Processes*, *134*, 1–3.

Hämäläinen, A., Dammhahn, M., Aujard, F., Eberle, M., Hardy, I., Kappeler, P. M., … Kraus, C. (2014). Senescence or selective disappearance? Age trajectories of body mass in wild and captive populations of a small-bodied primate. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1791).

Hansen, B. T., & Slagsvold, T. (2004). Early learning affects social dominance: interspecifically cross-fostered tits become subdominant. *Behavioral Ecology*, *15*(2), 262–268.

Hatch, M. I., & Smith, R. J. (2010). Repeatability of hematocrits and body mass of gray catbirds. *Journal of Field Ornithology*, *81*(1), 64–70.

Hauser, M. D. (1999). Perseveration, inhibition and the prefrontal cortex: a new look. *Current Opinion in Neurobiology*, *9*(2), 214–222.

Healy, S. D. (2012). Animal cognition: The trade-off to being smart. *Current Biology*, *22*(19), R840–R841.

Healy, S. D., De Kort, S. R., & Clayton, N. S. (2005). The hippocampus, spatial memory and food hoarding: A puzzle revisited. *Trends in Ecology and Evolution*, *20*(1), 17–22.

Henke-von der Malsburg, J., & Fichtel, C. (2018). Are generalists more innovative than specialists? A comparison of innovative abilities in two wild sympatric mouse lemur species. *Royal Society Open Science*, *5*(8), 180480.

Herrmann, E., & Call, J. (2012). Are there geniuses among the apes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2753–2761.

Herrmann, E., Call, J., Hernández-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.

Herrmann, E., Hernández-Lloreda, M. V., Call, J., Hare, B., & Tomasello, M. (2010). The structure of individual differences in the cognitive abilities of children and chimpanzees. *Psychological Science*, *21*(1), 102–110.

Hollis, K. L., Cadieux, E. L., & Colbert, M. M. (1989). The biological function of Pavlovian conditioning: A mechanism for mating success in the blue gourami (Trichogaster trichopterus). *Journal of Comparative Psychology*, *103*(2), 115–121.

Hopkins, W. D., Russell, J. L., & Schaeffer, J. (2014). Chimpanzee intelligence is heritable. *Current Biology*, *24*(14), 1649–1652.

Hõrak, P., Saks, L., Ots, I., & Kollist, H. (2002). Repeatability of condition indices in captive greenfinches (Carduelis chloris). *Canadian Journal of Zoology*, *80*(4), 636–643.

Houston, A., & McFarland, D. (1976). On the measurement of motivational variables. *Animal Behaviour*, *24*(2), 459–475.

Huber, L. (2017). Where is the evidence for general intelligence in nonhuman animals? *Behavioral and Brain Sciences*, *40*, e206.

Huchard, E., Canale, C. I., le Gros, C., Perret, M., Henry, P. Y., & Kappeler, P. M. (2012). Convenience polyandry or convenience polygyny? costly sex under female control in a promiscuous primate. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1732), 1371–1379.

Huebner, F., & Fichtel, C. (2015). Innovation and behavioral flexibility in wild redfronted lemurs (Eulemur rufifrons). *Animal Cognition*, *18*(3), 777–787.

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.

Hulthén, K., Chapman, B. B., Nilsson, P. A., Hansson, L.-A., Skov, C., Brodersen, J., … Brönmark, C. (2017). A predation cost to bold fish in the wild. *Scientific Reports*, *7*(1), 1239.

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology*. Oxford: 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.

Izquierdo, A., & Jentsch, J. D. (2012). Reversal learning as a measure of impulsive and compulsive behavior in addictions. *Psychopharmacology*, *219*(2), 607–620.

Jacobs, I. (2018). String Pulling. In J. Vonk & T. Shackelford (Eds.), *Encyclopedia of Animal Cognition and Behavior* (pp. 1–8). Springer International Publishing, Cham.

Jaumann, S., Scudelari, R., & Naug, D. (2013). Energetic cost of learning and memory can cause cognitive impairment in honeybees. *Biology Letters*, *9*(4), 20130149.

Johnstone, C. P., Reina, R. D., & Lill, A. (2012). Interpreting indices of physiological stress in free-living vertebrates. *Journal of Comparative Physiology B*, *182*(7), 861–879.

Joly, M., Ammersdörfer, S., Schmidtke, D., & Zimmermann, E. (2014). Touchscreen-based cognitive tasks reveal age-related impairment in a primate aging model, the grey mouse lemur (Microcebus murinus). *PLoS ONE*, *9*(10).

Jorgenson, J. T., Festa-Bianchet, M., Gaillard, J.-M., & Wishart, W. D. (1997). Effects of age, sex, disease, and density on survival of bighorn sheep. *Ecology*, *78*(4), 1019.

Kappeler, P. M., & Fichtel, C. (2012). A 15-year perspective on the social organization and life history of sifaka in Kirindy forest. In *Long-term field studies of primates* (pp. 101–121). Berlin, Heidelberg: Springer Berlin Heidelberg.

Kappeler, P. M., & Rasoloarison, R. M. (2003). Microcebus, mouse lemurs, Tsidy. In S. M. Goodman & J. P. Benstead (Eds.), *The Natural History of Madagascar* (pp. 1310–1315). Chicago: University of Chicago Press.

Kassambara, A., & Kosinski, M. (2017). Survminer: Drawing survival curves using ‘ggplot2’. R package version 0.4.0.

Kawecki, T. J. (2010). Evolutionary ecology of learning: insights from fruit flies. *Population Ecology*, *52*(1), 15–25.

Keagy, J., Savard, J. F., & Borgia, G. (2009). Male satin bowerbird problem-solving ability predicts mating success. *Animal Behaviour*, *78*(4), 809–817.

Keagy, J., Savard, J. F., & Borgia, G. (2011). Complex relationship between multiple measures of cognitive ability and male mating success in satin bowerbirds, Ptilonorhynchus violaceus. *Animal Behaviour*, *81*(5), 1063–1070.

Keagy, J., Savard, J. F., & Borgia, G. (2012). Cognitive ability and the evolution of multiple behavioral display traits. *Behavioral Ecology*, *23*(2), 448–456.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., … Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, *157*(3), 245–261.

Kittler, K., Kappeler, P. M., & Fichtel, C. (2018). Instrumental problem-solving abilities in three lemur species (Microcebus murinus, Varecia variegata, and Lemur catta). *Journal of Comparative Psychology*, *132*(3), 306–314.

Kolata, S., Light, K., & Matzel, L. D. (2008). Domain-specific and domain-general learning factors are expressed in genetically heterogeneous CD-1 mice. *Intelligence*, *36*(6), 619–629.

Kolm, N. (2014). Measuring variation in cognition can be done, but it requires hard empirical work: a comment on Rowe and Healy. *Behavioral Ecology*, *25*(6), 1296–1297.

Korte, S. M., Koolhaas, J. M., Wingfield, J. C., & McEwen, B. S. (2005). The Darwinian concept of stress: benefits of allostasis and costs of allostatic load and the trade-offs in health and disease. *Neuroscience & Biobehavioral Reviews*, *29*(1), 3–38.

Kotrschal, A., Rogell, B., Bundsen, A., Svensson, B., Zajitschek, S., Brännström, I., … Kolm, N. (2013). Artificial selection on relative brain size in the guppy reveals costs and benefits of evolving a larger brain. *Current Biology*, *23*(2), 168–171.

Kraus, C., Eberle, M., & Kappeler, P. M. (2008). The costs of risky male behaviour: Sex differences in seasonal survival in a small sexually monomorphic primate. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1643), 1635–1644.

Laland, K. N., & Reader, S. M. (2003). Animal innovation: An introduction. In S. M. Reader & K. N. Laland (Eds.), *Animal Innovation* (pp. 3–35).

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159.

Laughlin, S. B. (2001). Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*, *11*(4), 475–480.

Laughlin, S. B., de Ruyter van Steveninck, R. R., & Anderson, J. C. (1998). The metabolic cost of neural information. *Nature Neuroscience*, *1*(1), 36–41.

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

Lefebvre, L., & Sol, D. (2008). Brains, lifestyles and cognition: Are there general trends? *Brain, Behavior and Evolution*, *72*(2), 135–144.

Lermite, F., Peneaux, C., & Griffin, A. S. (2017). Personality and problem-solving in common mynas (Acridotheres tristis). *Behavioural Processes*, *134*, 87–94.

Light, K. R., Kolata, S., Wass, C., Denman-Brice, A., Zagalsky, R., & Matzel, L. D. (2010). Working memory training promotes general cognitive abilities in genetically heterogeneous mice. *Current Biology*, *20*(8), 777–782.

Lindström, J. (1999). Early development and fitness in birds and mammals. *Trends in Ecology & Evolution*, *14*(9), 343–348.

Locurto, C., Fortin, E., & Sullivan, R. (2003). The structure of individual differences in heterogeneous stock mice across problem types and motivational systems. *Genes, Brain and Behavior*, *2*(1), 40–55.

Lonnstedt, O. M., McCormick, M. I., Meekan, M. G., Ferrari, M. C. O., & Chivers, D. P. (2012). Learn and live: predator experience and feeding history determines prey behaviour and survival. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1736), 2091–2098.

Lührs, M. L., Dammhahn, M., Kappeler, P. M., & Fichtel, C. (2009). Spatial memory in the grey mouse lemur (Microcebus murinus). *Animal Cognition*, *12*(4), 599–609.

MacLean, E. L. (2016). Unraveling the evolution of uniquely human cognition. *Proceedings of the National Academy of Sciences*, *113*(23), 6348–6354.

MacLean, E. L., Barrickman, N. L., Johnson, E. M., & Wall, C. E. (2009). Sociality, ecology, and relative brain size in lemurs. *Journal of Human Evolution*, *56*(5), 471–478.

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*, *111*(20), E2140–E2148.

MacLean, E. L., Matthews, L. J., Hare, B. A., Nunn, C. L., Anderson, R. C., Aureli, F., … Wobber, V. (2012). How does cognition evolve? Phylogenetic comparative psychology. *Animal Cognition*, *15*(2), 223–238.

Maclean, E. L., Merritt, D. J., & Brannon, E. M. (2008). Social complexity predicts transitive reasoning in prosimian primates. *Animal Behaviour*, *76*(2), 479–486.

Madden, J. R., Langley, E. J. G., Whiteside, M. A., Beardsworth, C. E., & van Horik, J. O. (2018). The quick are the dead: pheasants that are slow to reverse a learned association survive for longer in the wild. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756), 20170297.

Maille, A., & Schradin, C. (2016). Survival is linked with reaction time and spatial memory in African striped mice. *Biology Letters*, *12*(8).

Matsuzawa, T. (2008). Primate foundations of human intelligence: A view of tool use in nonhuman primates and Fossil Hominids. In *Primate Origins of Human Cognition and Behavior* (pp. 3–25). Tokyo: Springer Japan.

Matzel, L. D., Han, Y. R., Grossman, H., Karnik, M. S., Patel, D., Scott, N., … Gandhi, C. C. (2003). Individual differences in the expression of a “general” learning ability in mice. *The Journal of Neuroscience*, *23*(16), 6423–6433.

Mazza, V., Eccard, J. A., Zaccaroni, M., Jacob, J., & Dammhahn, M. (2018). The fast and the flexible: cognitive style drives individual variation in cognition in a small mammal. *Animal Behaviour*, *137*, 119–132.

Mery, F. (2013). Natural variation in learning and memory. *Current Opinion in Neurobiology*, *23*(1), 52–56.

Mery, F., & Kawecki, T. J. (2003). A fitness cost of learning ability in Drosophila melanogaster. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1532), 2465–2469.

Milton, K. (1981). Distribution patterns of tropical plant foods as an evolutionary stimulus to primate mental development. *American Anthropologist*, *83*(3), 534–548.

Minter, R., Keagy, J., & Tinghitella, R. M. (2017). The relationship between male sexual signals, cognitive performance, and mating success in stickleback fish. *Ecology and Evolution*, *7*(15), 5621–5631.

Morand-Ferron, J. (2017). Why learn? The adaptive value of associative learning in wild populations. *Current Opinion in Behavioral Sciences*, *16*, 73–79.

Morand-Ferron, J., Cole, E. F., & Quinn, J. L. (2016). Studying the evolutionary ecology of cognition in the wild: A review of practical and conceptual challenges. *Biological Reviews*, *91*(2), 367–389.

Morand-Ferron, J., Cole, E. F., Rawles, J. E. C., & Quinn, J. L. (2011). Who are the innovators? A field experiment with 2 passerine species. *Behavioral Ecology*, *22*(6), 1241–1248.

Morand-Ferron, J., Hamblin, S., Cole, E. F., Aplin, L. M., & Quinn, J. L. (2015). Taking the operant paradigm into the field: Associative learning in wild great tits. *PLoS ONE*, *10*(8), 1–16.

Morand-Ferron, J., & Quinn, J. L. (2015). The evolution of cognition in natural populations. *Trends in Cognitive Sciences*, *19*(5), 235–237.

Murray, D. L. (2002). Differential body condition and vulnerability to predation in snowshoe hares. *Journal of Animal Ecology*, *71*(4), 614–625.

Nakagawa, S., & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews*, *85*(4), 935–956.

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research & Evaluation*, *9*(11).

Ots, I., MurumAgi, A., & Horak, P. (1998). Haematological health state indices of reproducing great tits: methodology and sources of natural variation. *Functional Ecology*, *12*(4), 700–707.

Øverli, Ø., Sørensen, C., Pulman, K. G. T., Pottinger, T. G., Korzan, W., Summers, C. H., & Nilsson, G. E. (2007). Evolutionary background for stress-coping styles: Relationships between physiological, behavioral, and cognitive traits in non-mammalian vertebrates. *Neuroscience & Biobehavioral Reviews*, *31*(3), 396–412.

Ozgul, A., Armitage, K. B., Blumstein, D. T., & Oli, M. K. (2006). Spatiotemporal variation in survival rates: implications for population dynamics of yellow-bellied marmots. *Ecology*, *87*(4), 1027–1037.

Plomin, R. (2001). The genetics of G in human and mouse. *Nature Reviews Neuroscience*, *2*(2), 136–141.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: five special findings. *Molecular Psychiatry*, *20*(1), 98–108.

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Sciences*, *7*(4), 157–160.

Pravosudov, V. V., & Clayton, N. S. (2002). A test of the adaptive specialization hypothesis: Population differences in caching, memory, and the hippocampus in black-capped chickadees (Poecile atricapilla). *Behavioral Neuroscience*, *116*(4), 515–522.

Pravosudov, V. V., & Roth, T. C. (2013). Cognitive ecology of food hoarding: The evolution of spatial memory and the hippocampus. *Annual Review of Ecology, Evolution, and Systematics*, *44*(1), 173–193.

Preiszner, B., Papp, S., Pipoly, I., Seress, G., Vincze, E., Liker, A., & Bokony, V. (2017). Problem-solving performance and reproductive success of great tits in urban and forest habitats. *Animal Cognition*, *20*(1), 53–63.

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.

R Core Team. (2017). R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Radespiel, U. (2006). Ecological diversity and seasonal adaptations of mouse lemurs (Microcebus spp.). In L. Gould & M. L. Sauther (Eds.), *Lemurs: Ecology and adaptation* (pp. 211–234). Boston, MA: Springer US.

Rahlfs, M., & Fichtel, C. (2010). Anti-predator behaviour in a nocturnal primate, the grey mouse lemur (Microcebus murinus). *Ethology*, *116*(5), 429–439.

Rahmani, H., Hoffmann, D., Walzer, A., & Schausberger, P. (2009). Adaptive learning in the foraging behavior of the predatory mite Phytoseiulus persimilis. *Behavioral Ecology*, *20*(5), 946–950.

Raine, N. E., & Chittka, L. (2008). The correlation of learning speed and natural foraging success in bumble-bees. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1636), 803–

808.

Rakotoniaina, J. H., Kappeler, P. M., Kaesler, E., Hämäläinen, A. M., Kirschbaum, C., & Kraus, C. (2017). Hair cortisol concentrations correlate negatively with survival in a wild primate population. *BMC Ecology*, *17*(1), 1–13.

Rakotoniaina, J. H., Kappeler, P. M., Ravoniarimbinina, P., Pechouskova, E., Hämäläinen, A. M., Grass, J., … Kraus, C. (2016). Does habitat disturbance affect stress, body condition and parasitism in two sympatric lemurs? *Conservation Physiology*, *4*(1), cow034.

Rasoazanabary, E. (2006). Male and female activity patterns in Microcebus murinus during the dry season at Kirindy Forest, Western Madagascar. *International Journal of Primatology*, *27*(2), 437–464.

Rasoloarison, R. M., Goodman, S. M., & Ganzhorn, J. U. (2000). A taxonomic revision of mouse lemurs (Microcebus) occurring in the western portion of Madagascar. *International Journal of Primatology*, *21*(6), 963–1019.

Rasoloarison, R. M., Rasolonandrasana, B., Ganzhorn, J. U., & Goodman, S. M. (1995). Predation on vertebrates in the Kirindy Forest, western Madagascar. *Ecotropica*, (1), 59–65.

Reader, S. (2003). Innovation and social learning: individual variation and brain evolution. *Animal Biology*, *53*(2), 147–158.

Reader, S. M., Hager, Y., & Laland, K. N. (2011). The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 1017–1027.

Reader, S. M., & Laland, K. N. (2001). Primate innovation: Sex, age and social rank differences. *International Journal of Primatology*, *22*(5), 787–805.

Reader, S. M., & Laland, K. N. (2002). Social intelligence, innovation, and enhanced brain size in primates. *Proceedings of the National Academy of Sciences*, *99*(7), 4436–4441.

Reader, S. M., & Laland, K. N. (2003). *Animal Innovation*. Oxford University Press.

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318.

Reeve, C. L., Lyerly, J. E., & Peach, H. (2013). Adolescent intelligence and socio-economic wealth independently predict adult marital and reproductive behavior. *Intelligence*, *41*(5), 358–365.

Rensing, S. (1999). Immobilization and anesthesia of nonhuman primates. *Primate Report*, *55*, 33–38.

Roff, D. A. (1992). *The evolution of life histories: theory and analysis*. New York: Chapman & Hall.

Roff, D. A. (2002). *Life history evolution*. Sunderland, Mass.: Sinauer Associates, Inc.

Romero, L. M., Dickens, M. J., & Cyr, N. E. (2009). The reactive scope model — A new model integrating homeostasis, allostasis, and stress. *Hormones and Behavior*, *55*(3), 375–389.

Roth, G., & Dicke, U. (2005). Evolution of the brain and intelligence. *Trends in Cognitive Sciences*, *9*(5), 250–257.

Roth, T. C., LaDage, L. D., Freas, C. A., & Pravosudov, V. V. (2012). Variation in memory and the hippocampus across populations from different climates: a common garden approach. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1727), 402–410.

Roth, T. C., LaDage, L. D., & Pravosudov, V. V. (2010). Learning capabilities enhanced in harsh environments: A common garden approach. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1697), 3187–3193.

Rowe, C., & Healy, S. D. (2014). Measuring variation in cognition. *Behavioral Ecology*, *25*(6), 1287–1292.

Sauce, B., Bendrath, S., Herzfeld, M., Siegel, D., Style, C., Rab, S., … Matzel, L. D. (2018). The impact of environmental interventions among mouse siblings on the heritability and malleability of general cognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1756), 20170289.

Scheid, C., & Bugnyar, T. (2008). Short-term observational spatial memory in Jackdaws (Corvus monedula) and Ravens (Corvus corax). *Animal Cognition*, *11*(4), 691–698.

Scheumann, M., Rabesandratana, A., & Zimmermann, E. (2007). Predation, communication, and cognition in lemurs. In *Primate Anti-Predator Strategies* (pp. 100–126). Boston, MA: Springer US.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113.

Schliehe-Diecks, S., Eberle, M., & Kappeler, P. M. (2012). Walk the line-dispersal movements of gray mouse lemurs (Microcebus murinus). *Behavioral Ecology and Sociobiology*, *66*(8), 1175–1185.

Schmid, J. (1999). Sex-specific differences in activity patterns and fattening in the gray mouse lemur (Microcebus murinus) in Madagascar. *Journal of Mammalogy*, *80*(3), 749–757.

Schmid, J., & Kappeler, P. (1998). Fluctuating sexual dimorphism and differential hibernation by sex in a primate , the gray mouse lemur (Microcebus murinus). *Behavioral Ecology and Sociobiology*, *43*, 125–132.

Schulte-Hostedde, A. I., Zinner, B., Millar, J. S., & Hickling, G. J. (2005). Restitution of mass-size residuals: validating body condition indices. *Ecology*, *86*(1), 155–163.

Shatz, S. M. (2008). IQ and fertility: A cross-national study. *Intelligence*, *36*(2), 109–111.

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.

Shaw, R. C., Boogert, N. J., Clayton, N. S., & Burns, K. C. (2015). Wild psychometrics: Evidence for “general” cognitive performance in wild New Zealand robins, Petroica longipes. *Animal Behaviour*, *109*, 101–111.

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.

Shettleworth, S. J. (1990). Spatial memory in food-storing birds. *Philosophical Transactions: Biological Sciences*. Royal Society.

Shettleworth, S. J. (2010). *Cognition, evolution, and behavior* (2nd ed.). Oxford: Oxford University Press.

Shettleworth, S. J. (2012). Modularity, comparative cognition and human uniqueness. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2794–2802.

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*(1603), 2762–2772.

Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century-Crofts.

Slagsvold, T., & Wiebe, K. L. (2007). Learning the ecological niche. *Proceedings. Biological Sciences*, *274*(1606), 19–23.

Smith, C., Philips, A., & Reichard, M. (2015). Cognitive ability is heritable and predicts the success of an alternative mating tactic. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1809).

Snell-Rood, E. C., Davidowitz, G., & Papaj, D. R. (2011). Reproductive tradeoffs of learning in a butterfly. *Behavioral Ecology*, *22*(2), 291–302.

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*, *102*(15), 5460–5465.

Sol, D., Griffin, A. S., & Bartomeus, I. (2012). Consumer and motor innovation in the common myna: The role of motivation and emotional responses. *Animal Behaviour*, *83*(1), 179–188.

Sol, D., Lefebvre, L., & Rodríguez-Teijeiro, J. D. (2005). Brain size, innovative propensity and migratory behaviour in temperate Palaearctic birds. *Proceedings. Biological Sciences*, *272*(1571), 1433–1441.

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.

Sol, D., Timmermans, S., & Lefebvre, L. (2002). Behavioural flexibility and invasion success in birds. *Animal Behaviour*, *63*(3), 495–502.

Stalder, T., & Kirschbaum, C. (2012). Analysis of cortisol in hair - State of the art and future directions. *Brain, Behavior, and Immunity*, *26*(7), 1019–1029.

Stearns, S. C. (1989). Trade-offs in life-history evolution. *Functional Ecology*, *3*(3), 259.

Stearns, S. C. (1992). *The evolution of life histories*. New York: Oxford University Press.

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.

Strenze, T. (2006). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* , *35*(5), 401-426.

Ten Cate, C. (2014). Towards fruitful interaction between behavioral ecology and cognitive science: A comment on Rowe and Healy. *Behavioral Ecology*, *25*(6), 1295–1296.

Therneau, T. (2015). A package for survival analysis in S. R package version 2.38.

Thornton, A., Isden, J., & Madden, J. R. (2014). Toward wild psychometrics: Linking individual cognitive differences to fitness. *Behavioral Ecology*, *25*(6), 1299–1301.

Thornton, A., & Lukas, D. (2012). Individual variation in cognitive performance: Developmental and evolutionary perspectives. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2773–2783.

Thornton, A., & Samson, J. (2012). Innovative problem solving in wild meerkats. *Animal Behaviour*, *83*(6), 1459–1468.

Tinbergen, J. M., & Boerlijst, M. C. (1990). Nestling weight and survival in individual great tits (Parus major). *The Journal of Animal Ecology*, *59*(3), 1113.

van de Waal, E., Borgeaud, C., & Whiten, A. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science*, *340*(6131), 483–485.

van de Waal, E., & Bshary, R. (2010). Contact with human facilities appears to enhance technical skills in wild vervet monkeys (Chlorocebus aethiops). *Folia Primatologica*, *81*(5), 282–291.

van Horik, J. O., Clayton, N. S., & Emery, N. J. (2012). Convergent evolution of cognition in corvids, apes and other animals. *The Oxford Handbook of Comparative Evolutionary Psychology*, 80–101.

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).

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).

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.

van Horik, J. O., & Lea, S. E. G. (2017). Disentangling learning from knowing: Does associative learning ability underlie performances on cognitive test batteries? *Behavioral and Brain Sciences*, *40*, e220.

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.

Völter, C., 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*, *373*(1756).

Vuarin, P., Dammhahn, M., & Henry, P. Y. (2013). Individual flexibility in energy saving: Body size and condition constrain torpor use. *Functional Ecology*, *27*(3), 793–799.

Wahlsten, D. (1972). Genetic experiments with animal learning: A critical review. *Behavioral Biology*, *7*(2), 143–182.

Walker, B. G., Boersma, P. D., & Wingfield, J. C. (2005). Field endocrinology and conservation biology. *Integrative and Comparative Biology*, *45*(1), 12–18.

Wass, C., Denman-Brice, A., Rios, C., Light, K. R., Kolata, S., Smith, A. M., & Matzel, L. D. (2012). Covariation of learning and “reasoning” abilities in mice: evolutionary conservation of the operations of intelligence. *Journal of Experimental Psychology. Animal Behavior Processes*, *38*(2), 109–124.

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.

Wilson, A. J., & Nussey, D. H. (2010). What is individual quality? An evolutionary perspective. *Trends in Ecology & Evolution*, *25*(4), 207–214.

Witter, M. S., & Cuthill, I. C. (1993). The ecological costs of avian fat storage. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *340*(1291), 73–92.

Würbel, H. (2001). Ideal homes? Housing effects on rodent brain and behaviour. *Trends in Neurosciences*, *24*(4), 207–211.

Ydenberg, R. C. (1994). The behavioral ecology of provisioning in birds. *Écoscience*, *1*(1), 1–14.

Yoder, A. D. (2007). Lemurs. *Current Biology*, *17*(20), R866-8.

Zimmermann, E., Radespiel, U., Mestre-Francés, N., & Verdier, J.-M. (2016). Life history variation in mouse lemurs (Microcebus murinus, M. lehilahytsara): the effect of environmental and phylogenetic determinants. In S. M. Lehman, U. Radespiel, & E. Zimmermann (Eds.), *The Dwarf and Mouse Lemurs of Madagascar* (pp. 174–194). Cambridge: Cambridge University Press.

Acknowledgements

There are so many people who have helped and accompanied me during this PhD and contributed to its completion. First, I want to thank Claudia Fichtel and Peter Kappeler for the long way on which you accompanied me. I am very grateful that you took me with you to Kirindy for the master field course back in 2012. This first 6 weeks in Madagascar broadened my horizons so intensively and was the beginning of all that has led to this PhD thesis and so much more. Peter, I thank you for the unique opportunity of being actively involved in forming the ideas and framework for my PhD project, so that I could work on a topic that really excites me. I am deeply thankful to both of you for your scientific guidance and experience on the way to realize and finalize this PhD project. I especially thank Claudia, my supervisor, for her constant support during all this time, be it scientifically or morally during all the ups and downs on the way from the first internship and the master thesis to the end of the PhD.

I am very thankful to my thesis committee members Lars Penke and Oliver Schülke for their time, valuable input and advice. Furthermore, I want to thank Julia Fischer and Margarete Boos for being in my examination board.

I thank Beate Kraus and Lynne Werner for their help with the video analysis and inter-observer reliability.

I want to thank Joah Madden for inviting me to an exciting workshop on the causes and consequences of individual differences in cognitive abilities, his helpful advices and his understanding of the challenges involved in testing 96 wild living mouse lemurs.

Furthermore, I thank Roger Mundry for his great, intensive, mind-opening statistic workshop and his uniquely comprehensive, funny and friendly way to teach which deeply impressed me.

I would also like to thank all of my past and present colleagues and friends from the Behavioural ecology/ Sociobiology unit. Thank you for your help and support, without you these last years would not have been half as much fun! I thank Christina Glaschke for her work on the parentage analysis for my project. Moreover, I especially thank my long-time office mate and friend Klara Dinter, who I also thank a lot for proof reading parts of this thesis! Thanks!

My deepest gratitude goes to all members of l’équipe de Kirindy for the unique time in the field: Leon, Rodin, Pata, Bary, Adrien, Alain, Mamy, Remy, Bruno and Jipa. Thank you for the nice conservations, guitar and singing evenings, the best beans I ever ate, and a lot of “tsanga tsanga” and shared hard work (be it eating hot soup for lunch in what felt like 40 degrees outside temperature or the rush hour setting of 160 traps before the sun sets). I especially thank Bruno with whom I really enjoyed working, spending peaceful hours sitting next to each other handling mouse lemurs or cleaning traps. Misaotra betsaka! Without you all it wouldn’t have been the same.

Another special thanks goes to my companions in the field who spent so much time with me in the forest and on the bi-weekly weekends in Morondava. Thanks especially to:
Bako Rasolofoniaina, with whom I loved to go hunting for precious cauliflower and the best pineapples at the Morondava market.
Anna Sperber, with whom I spent funnily desperate weekend nights in the Morondava townhouse, sitting in the candlelight and waiting for the end of the electricity cut-off, begging for a tiny bit of electricity for the much sighed-for skype calls with the loved ones at home.
And to Louise Peckre and Alessio Anania, who let me view the forest again with the eyes of the first-time visitor when they joined the Kirindy team with their contagious passion for all living beings around us.

Wholeheartedly, I want to thank my friends and family. Thank you for your endless support and love and for having you in my life. I thank my father for all his help and practical support with no matter what, be it escorting 100 microtransponders to Kirindy Forest or building a touch-screen testing device for the mouse lemurs. I thank my mother for all her love and for not worrying too much when I was away. I thank my sister for always saying the right words to cheer me up. And finally, I thank Hendrik for always being at my side, be it in thought during the times in the forest, via skype on the city weekends in Morondava or with all possible help and support in Göttingen, especially during the ups and downs in the last months of my PhD. Thank you!

Curriculum vitae

Franziska Hübner

Personal information

Date of birth: 23.01.1988
Place of birth: Munich
Nationality: German

Education

2015- 2019	PhD student at Georg-August-University Göttingen and German Primate Center (DPZ) Thesis: “The link between cognition and fitness in wild grey mouse lemurs (*Microcebus murinus*)”
2011-2014	Master of science in Developmental, Neural and Behavioural Biology at Georg-August-University Göttingen Thesis: "Innovation and behavioural flexibility in wild redfronted lemurs (*Eulemur rufifrons*)"
2008-2011	Bachelor of science in Biology at Georg-August-University Göttingen Thesis: "Male dispersal and behaviour of four-year-old stallions in a herd of semi-free-ranging Exmoor ponies (*Equus ferus caballus*) at Langeland, Denmark"

Research experience

2015-2017	Data collection for the PhD thesis, Kirindy forest, Madagascar (in total 11 months)
2014	Pilot study for the PhD thesis, Kirindy forest, Madagascar (3 months)
2014	Research assistant in the “Behavioural Ecology and Sociobiology Unit” at the German Primate center
3/2012, 3/2013, 3/2014	Student assistant in the practical course “Animal physiology” for Bachelor students of Biology at Georg-August-University Göttingen
2013	Data collection for the Master thesis, Kirindy forest, Madagascar (3 months)
2011	Data collection for the Bachelor thesis, Langeland, Denmark (2 months)

Grants and awards

Research grant of the Ethologische Gesellschaft, for the master thesis, 2013
Poster award, Conference of the Gesellschaft für Primatologie, Leipzig, 2015
Poster award, Conference of the Gesellschaft für Primatologie, Zürich, 2017

Honorary office

PhD student representative of the German Primate Center, 2018

Conference contributions

Hübner F & Fichtel C. Innovation and behavioural flexibility in wild redfronted lemurs (*Eulemur rufifrons*). Poster, Conference of the Gesellschaft für Primatologie (GfP), Leipzig 2015

Hübner F, Fichtel C & Kappeler PM. Testing the link between cognition and fitness in a small, free-living primate, the grey mouse lemur (*Microcebus murinus*). Poster, Meeting of the Ethological Society, Göttingen, 2016

Hübner F, Fichtel C & Kappeler PM. Does it pay to be smart? The link between cognition and fitness in the grey mouse lemur (*Microcebus murinus*). Oral presentation, University of Antananarivo, 2016

Hübner F, Fichtel C & Kappeler PM. The link between cognition and fitness in a small, free-living primate, the grey mouse lemur (*Microcebus murinus*). Poster, Conference of the Gesellschaft für Primatologie (GfP), Zürich, 2017

Hübner F, Fichtel C & Kappeler PM. Does it pay to be smart? The link between cognition and fitness in the grey mouse lemur (*Microcebus murinus*). Oral presentation, Wissenschaftliches Kolleg-Studienstiftung des deutschen Volkes, Göttingen, 2017

Hübner F, Fichtel C & Kappeler PM. Individual differences in cognitive abilities in a small, wild primate, the grey mouse lemur (*Microcebus murinus*). Poster, International Ethological Conference (IEC) and Summer Meeting of the Association for the Study of Animal Behaviour (ASAB), Estoril, 2017

Hübner F, Fichtel C & Kappeler PM. Does it pay to be smart? Cognition and fitness in wild grey mouse lemurs, *Microcebus murinus.* Oral presentation, Workshop "Causes and consequences of individual differences in cognitive abilities", Exeter, 2017

Hübner F, Fichtel C.& Kappeler PM. Does it pay to be smart? The link between cognition and fitness in the grey mouse lemur (*Microcebus murinus*). Poster, Conference of the Research Training Group 2070, Göttingen, 2018

Publications

Hübner F, Fichtel C, Kappeler PM (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* 373: 20170295.

Cauchoix M, Chow PKY, van Horik JO, Atance CM, Barbeau EJ, Barragan-Jason G, Bize P, Boussard A, Buechel SD, Cabirol A, Cauchard L, Claidière N, Dalesman S, Devaud JM, Didic M, Doligez B, Fagot J, Fichtel C, Henke-von der Malsburg J, Hermer E, Huber L, Hübner F, Kappeler PM, Klein S, Langbein J, Langley EJG, Lea SEG, Lihoreau M, Lovlie H, Matzel LD, Nakagawa S, Nawroth C, Oesterwind S, Sauce B, Smith EA, Sorato E, Tebbich S, Wallis LJ, Whiteside MA, Wilkinson A, Chaine AS & Morand-Ferron J (2018). The repeatability of cognitive performance: a meta-analysis. *Philosophical Transactions of the Royal Society B* 373: 20170281.

Hübner F & Fichtel C (2015) Innovation and behavioral flexibility in redfronted lemurs (*Eulemur rufifrons*). *Animal Cognition* 18, 777-787.

Kraus C, van Waveren C & Hübner F (2014) Distractible dogs, constant cats? A test of the distraction hypothesis in two domestic species. *Animal Behaviour* 93, 173-181.

Schnoell AV, Hübner F, Kappeler PM & Fichtel C (2014) Manual lateralization in wild redfronted lemurs (*Eulemur rufifrons*) during spontaneous actions and in an experimental task. *American Journal of Physical Anthropology* 153, 61-67.

Declaration

I hereby declare that all parts of my thesis with the title "Cognition in the wild -Individual differences in cognitive abilities and their link with fitness in a wild primate, the grey mouse lemur (*Microcebus murinus*)" were written by myself. Assistance of third parties was only accepted if scientifically justifiable and acceptable in regards to the examination regulations. Contributions to the individual chapters are indicated and all sources have been quoted.

Göttingen, 24th of April, 2019