Title

Developmental Pathways of Early-Life Adversity: Behavioral, Physiological, and Life-History Consequences in Wild White-Faced Capuchin Monkeys (Cebus imitator)

ISBN

9798190948929

Author

Mensing, Ashley Nicole

Publication Date

2026-09-16

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Developmental Pathways of Early-Life Adversity: Behavioral, Physiological, and Life-History Consequences in Wild White-Faced Capuchin Monkeys (*Cebus imitator*)

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Anthropology

by

Ashley Mensing

2026

ABSTRACT OF THE DISSERTATION

Developmental Pathways of Early-Life Adversity: Behavioral, Physiological, and Life-History Consequences in Wild White-Faced Capuchin Monkeys (*Cebus imitator*)

by

Ashley Mensing

Doctor of Philosophy in Anthropology

University of California, Los Angeles, 2026

Professor Susan Perry, Chair

Early-life adversity (ELA) is hypothesized to shape adult phenotypes and fitness outcomes through developmental effects on physiological and behavioral systems, yet evidence from wild, long-lived primates remains limited. This dissertation investigates the relationships among ELA, hypothalamic-pituitary-adrenal (HPA) axis regulation, behavioral phenotypes, and life-history outcomes in white-faced capuchin monkeys (*Cebus imitator*) at Lomas Barbudal Biological Reserve in Guanacaste, Costa Rica, using over three decades of longitudinal data from the Lomas Barbudal Monkey Project.

Chapter 1 examines ELA effects on female mortality and reproductive trajectories across three developmental windows. Prenatal exposure to alpha male replacement increased mortality risk by 75%, while psychosocial adversity (alpha male replacements, maternal loss) accelerated reproductive timing among females who survived to the core reproductive period. Chapter 2 evaluates whether boldness and aggression covary into the integrated behavioral syndrome theory proposed as the mediator linking early developmental conditions to adult life-history

strategy. Boldness showed very low repeatability, aggression was target-specific rather than trait-like, and no population-level behavioral syndrome emerged, though individual-level heterogeneity in trait covariance was detected, with bolder individuals showing greater behavioral flexibility. Chapter 3 develops a rolling-window methodology for extracting biologically meaningful HPA axis phenotypes from field-collected fecal glucocorticoid samples. Tonic and reactive HPA axis phenotypes were uncorrelated, demonstrating that tonic glucocorticoid activity and acute stress responsiveness represent independent dimensions of endocrine function. Chapter 4 integrates these components to test whether ELA programs adult HPA axis function, which in turn shapes behavior and male life-history outcomes. Evidence for developmental programming of the HPA axis was limited; however, proactive aggression was associated with a higher probability of alpha male attainment, though this finding should be interpreted cautiously given the small number of attainment events. The hypothesized ELA-to-HPA axis-to-behavior-to-fitness mediation chain was not supported.

Collectively, these findings demonstrate that theoretical predictions derived from laboratory and human studies do not straightforwardly translate to wild, long-lived primates. The results highlight the importance of direct ecological constraints over predictive developmental programming and reveal that behavioral and physiological phenotypes operate as largely independent dimensions in shaping fitness outcomes.

The dissertation of Ashley Mensing is approved.

Jessica Lynch

Brooke Scelza

Jacinta Beehner

Tessa Solomon-Lane

Susan Perry, Committee Chair

University of California, Los Angeles
2026

For David Aníbal, who arrived near the end of this journey and reminded me that there are more important things than dissertations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

Chapter 1 of this dissertation “Early-Life Adversity Shapes Mortality and Reproductive Strategies in Wild Female Capuchin Monkeys” is a version of a manuscript submitted for publication to the American Journal of Primatology and currently under review. Susan Perry is a co-author on this manuscript and contributed to the conceptualization and design, data collection, funding, and oversight as PI.

Along the way, this dissertation was shaped by mentors, colleagues, friends, family, and an extraordinary community of humans and monkeys. Whatever value this work has is the product of their generosity.

I begin with Susan Perry. It is no exaggeration to say that Susan changed the course of my life. I remain grateful that she decided to take a chance on me in 2017 as an intern on the Lomas Barbudal Monkey Project, this came at a pivotal moment in my life, a life line out of a difficult situation – my life would today would be unrecognizable were it not for her invitation to move to Costa Rica and spend a year chasing monkeys through the jungle and learning to be a primatologist from her. Over the years, Susan has been generous with her data, her time, her mentorship, and often her dinner table. She taught me the value of rigor, skepticism, and careful scientific thinking. More than any specific result in this dissertation, I suspect that commitment to getting the science right will be the most enduring lesson I carry forward from her mentorship. She advocated for me when I needed support, stood by me during difficult periods, and repeatedly opened doors that changed the trajectory of my career.

I am deeply grateful to the rest of my dissertation committee. Brooke Scelza joined my intellectual life during my very first quarter at UCLA. After taking her seminar on Feminist Evolutionists, I knew she was someone I wanted to learn from. Brooke's comments were

invariably incisive, forcing me to think more carefully about socioecological explanations and alternative interpretations. She has a gift for identifying exactly the question that still needs answering.

Jessica Lynch brought perspectives that broadened both my science and my understanding of the responsibilities that accompany field research. Through her encouragement, I engaged more deeply with Spanish-language scholarship and became more attentive to the scientific communities of the countries in which we work. She was consistently supportive and generous with her time, and her mentorship challenged me to think more broadly and more carefully.

Tessa Solomon-Lane helped connect my work to developmental biology and reminded me that primatology is part of a much larger scientific conversation. She was remarkably generous with practical guidance, sharing analytical advice, R code, and examples of successful grant applications when I was still learning how to navigate graduate school. She was far more engaged than I ever expected and far more helpful than I had any right to ask and for that I am grateful.

Long before Jacinta Beehner became a committee member, collaborator, or mentor, she was an intellectual influence. In 2017, I downloaded a set of papers from a literature review she had prepared and read them in the forest on days when the monkeys could not be found. Those papers introduced me to field endocrinology and helped me see how questions from developmental psychology could be translated into a primate field setting. Years later, it felt fitting to collaborate with her on the El Niño project, learn from her directly, and ultimately have her serve on my committee. Jacinta welcomed me into her lab, shared equipment and expertise freely, and taught me an enormous amount about endocrinology and scientific pragmatism.

Before UCLA, I was fortunate to work with mentors who encouraged me to become a scientist long before I believed I could be one. I thank Yaakov Stern, Elaine Gazes, and Gil Rabinovici for their patience, generosity, and willingness to let an insatiably curious research assistant wander beyond the boundaries of her assigned responsibilities. They hired me to support their research, but they also allowed me to learn from it. I inserted myself into their science whenever possible, and they responded with encouragement rather than irritation (most of the time, outwardly). The lessons I learned from them shaped every stage of my career.

My first field season at Lomas Barbudal was shared with Natalia, Luz, David, Jamie, Fiona, Odd, Maggie, Amaia, Rupert, Alexa, and Avery. We woke before sunrise, rode into the forest every morning, returned home exhausted each evening, and gathered around the dinner table to debate science, life, and everything in between. We were not always harmonious. We argued, disagreed, and occasionally drove one another crazy. But we also supported one another, challenged one another intellectually, and built a sense of solidarity that remains one of my favorite memories of that time. Many of them have gone on to become scientists in their own right, which surprises me less every year.

Field research depends on people who show up day after day regardless of the heat, rain, insects, setbacks, or monkey-related frustrations. I am grateful to Adam Gaston for his effort and professionalism during the 2023-2024 field season, a particularly challenging one for me personally. Field research is sustained by countless people whose names rarely appear in publications. Marlon rescued us more times than I can count. I am especially grateful for the night we found ourselves stranded by a flooded river, separated from our vehicle and forced to hike for hours through the dark and rain until we finally reached a highway and found enough signal to make a phone call. When Marlon answered and came to get us, I do not think I had ever

been so relieved to see another human being. I am also grateful to the staff of the Lomas Barbudal Biological Reserve and to the owners of Pelón de la Bajura for their generosity and support. I also thank Seidy for cool water, grilled cheese sandwiches, and hospitality on impossibly hot afternoons.

Graduate school would have been immeasurably poorer without the friendships of Sophie Klitgaard, Sally Li, Kasey Schleper, Naima Tucker, and especially Tlaoli Anaya-Fuentes. Some of my favorite memories of UCLA involve conversations on trails winding through the mountains surrounding Los Angeles. Those conversations sustained me intellectually and personally through the many challenges of graduate school. Tlaoli deserves special thanks, my academic sister/fellow advisee. In 2017, during a visa trip to Nicaragua, I was caught in a rip current and being pulled farther from shore. When I finally managed to get my head above water, Tlaoli helped me recognize what was happening, reminded me to swim horizontally, and helped guide me to safety. Looking back, I realize that moment became an apt metaphor for much of graduate school. Time and again, she helped me recognize when I was in trouble and reminded me to keep moving.

I thank the National Science Foundation for support through a Doctoral Dissertation Research Improvement Grant, the UCLA Division of Graduate Education for support through the Graduate Dean's Scholar Award and Dissertation Year Fellowship, and the UCLA Department of Anthropology. I am especially grateful to Jason Throop and Erin Debenport, who rescued me during my first summer at UCLA when I failed to read the fine print of an award letter and found myself in an administrative predicament entirely of my own making.

To my grandmother: thank you for never missing an opportunity to tell me how proud you are. I am grateful beyond words. To my mother and father, thank you for your support and encouragement in the ways you were able to give it.

This dissertation would not have been possible without the generosity of colleagues who shared both expertise and resources. I am especially grateful to Irene Godoy and Colleen Gault, whose earlier work helped shape many of the questions that eventually became this dissertation. Long before these chapters existed, they shared their knowledge, experiences, and biological samples, and offered advice during the earliest stages of project development. Their willingness to help a much more junior researcher navigate the initial stages of dissertation planning had a lasting impact on the direction of this work.

I also owe a tremendous debt to Sofia Carrera, Teera Losch, and Walker Stanton at the University of Michigan. Field endocrinology often appears deceptively straightforward when described in a methods section. In reality, the process depends on a great deal of technical expertise that is rarely visible on the page. Sofia and Teera taught me the practical realities of running hormone assays, patiently answered countless questions, and helped me develop the skills necessary to generate the data that form the foundation of this dissertation.

I also thank Kotrina Kajokaite and Siavash Jalal of the UCLA Statistical Consulting Group. Many of the analyses in this dissertation benefitted from their expertise, creativity, and willingness to help me think through difficult methodological problems. They improved both the rigor of the work and my understanding of the statistical tools used to conduct it.

And I could not forget to thank Don Cohen. Many years were spent sandwiched on the couch between Don and Susan, each with our respective laptops spending hours at a time thinking through edge cases and testing the logic of every assumption that fed every query that

built every variable used in these analyses. The intellectual labor that he contributed towards this dissertation is significant. He inspired in me a love for coding and logic puzzles that surprised me and was always generous with his time and snacks. Finally, he provided an abundance of opportunities to practice my skills as a collaborator, to build on areas of agreement, learn to compromise, and concede (often after vigorous debate) when proven wrong.

My deepest gratitude belongs to my husband, David Amaury. Since we met in 2019, he has followed me across the country, left behind his own business and community, moved with me to California, joined me in Costa Rica, accompanied me to the Everglades for side projects that had absolutely nothing to do with this dissertation, and never once turned down an adventure. He has supported this work emotionally, materially, and practically in ways too numerous to count. In 2025, while I was five months pregnant and racing to complete the laboratory work that forms the foundation of Chapter 3, Amaury drove with me from Los Angeles to Ann Arbor and spent five weeks helping process more than 1,300 hormone samples. Together we condensed nearly two months of laboratory work into that short period of time. He performed extractions while I ran assays and we worked back-to-back 12-hour days while living in an extended-stay hotel that was, charitably speaking, not ideal. He has supported every stage of this project because he believes in me and I quite literally could not have done this without him.

Finally, I thank the capuchins, but especially those of the FL, FF, and MI groups. Thank you for sharing your personalities so generously, bringing me to some of the best and worst places, for your poops, and the countless hours you allowed me to spend in your company. Thank you for helping me become not only a scientist, but a more human human.

VITA
ASHLEY MENSING

EDUCATION

PhD in Anthropology (Biological Anthropology) — In progress
University of California, Los Angeles

MA in Developmental Psychology — May 2021
Columbia University, Teacher's College

BS in Interdisciplinary Social Sciences — December 2011
Florida State University

RESEARCH POSITIONS

Study Manager / Research Assistant — 2018–2021
Taub Institute, Cognitive Neuroscience Division, Columbia University (PI: Yaakov Stern)

Field Assistant / Lab Manager — 2017–2018
Lomas Barbudal Monkey Project, Costa Rica (PI: Susan Perry)

Behavioral Observation Team Member — 2016–2017
Oakland Zoo, Oakland, CA

Project Manager — 2015–2017
Memory and Aging Center, University of California, San Francisco

Clinical Research Coordinator — 2014–2015
Olympian Clinical Research, Tampa, FL

Program Supervisor — 2012–2014
Hilltop Research, St. Petersburg, FL

Research Assistant — 2008–2009
Department of Biology, Florida State University

TEACHING EXPERIENCE

Teaching Assistant — 2022 - 2026
Department of Anthropology, UCLA

PUBLICATIONS

Peer-Reviewed Journal Articles

Carrera, S., Godoy, I., Gault, C., **Mensing, A.**, Damm, J., Perry, S., & Beehner, J. (2025). Stress responsiveness in a wild primate predicts survival across an extreme El Niño drought. *Science Advances, 11*(4), eadq5020.

Gazes, Y., Lee, S., Fang, Z., **Mensing, A.**, Noofoory, D., Hidalgo Nazario, G., ... & Stern, Y. (2023). Effects of brain maintenance and cognitive reserve on age-related decline in three cognitive abilities. *The Journals of Gerontology: Series B*, gbad044.

Simon, S. S., Lee, S., Gu, Y., **Mensing, A.**, Noofoory, D., Nazario, G. M. H., ... & Stern, Y. (2022). Leisure activity engagement across adulthood predicts cognitive change after five years: Do gender and age matter? *Journal of the International Neuropsychological Society*, 1–12.

Sunderaraman, P., Gazes, Y., Ortiz, G., Langfield, C., **Mensing, A.**, Chapman, S., ... & Cosentino, S. (2022). Financial decision-making and self-awareness for financial decision-making is associated with white matter integrity in older adults. *Human Brain Mapping, 43*(5), 1630–1639.

Hagen, R. V., Knorr, D. A., Li, S., **Mensing, A.**, & Scelza, B. A. (2022). Staying alive enhances both women's and men's fitness. *Behavioral and Brain Sciences, 45*.

Gazes, Y., Lee, S., Sakhardande, J., **Mensing, A.**, Razlighi, Q., Ohkawa, A., Pleshkevich, M., Luo, L., & Habeck, C. (2020). fMRI-guided white matter connectivity in fluid and crystallized cognitive abilities. *NeuroImage, 215*.

La Joie, R., Bejanin, A., Fagan, A., Ayakta, N., Baker, S., Bourakova, V., Boxer, A., Cha, J., Karydas, A., Jerome, G., Maass, A., **Mensing, A.**, Miller, Z., O'Neil, J., Pham, J., Rosen, H., Tsai, R., Visani, A., Miller, B., Jagust, W., & Rabinovici, G. (2018). Associations between [18F]AV1451 Tau PET and CSF measures of tau pathology in a clinical sample. *Neurology, 90*(4).

Mensing, A. (2012). Play in wild, immature, white-faced capuchins. *Florida State Undergraduate Research Journal, 2*(1).

INVITED TALKS

2026 "Early-life adversity and life history correlates in wild, white-faced capuchins: physiological mechanisms and fitness outcomes.", *Behavior, Evolution, and Culture Speaker Series, University of California, Los Angeles,* Los Angeles, CA

CONFERENCE PRESENTATIONS

Mensing, A. (2026). Early-life adversity shapes mortality and reproductive strategies in wild female capuchin monkeys. Podium talk presented at the American Society of Primatologists Annual Conference, Arlington, VA.

Mensing, A. (2024). Confronting the serpent: Risk-taking and mortality perceptions in python pursuits. Poster presented at the American Anthropological Association Annual Conference, Tampa, FL.

Mensing, A. & Perry, S. (2023). Methods for measuring individual differences in stress responsivity following intergroup encounters in white-faced capuchins (*Cebus imitator*). Poster presented at Ecology, Biology, Health and Behavior of White-Faced Capuchins Conference.

Mensing, A. & Klitgard, S. (2023). Are there ethical risks to applying life-history theory to human behavioral variation? Poster presented at the Human Behavior and Evolution Society Annual Conference.

Mensing, A. & Perry, S. (2023). Extrinsic mortality and the limits of agency (or: A test of the harsh environment, fast life history hypothesis in capuchins). Talk presented at California Workshop on Evolutionary Social Sciences.

Mensing, A., Maurer, K., Das, N., Burns, E., & Stern, Y. (2020). Low socioeconomic status as a proxy for stress in the brain. Poster presented at the Cognitive Neuroscience Society Annual Conference.

Mensing, A., Gazes, Y., Gu, Y., & Stern, Y. (2019). Socioeconomic status predicts cognition across the lifespan. Poster presented at the International Neuropsychological Society Annual Conference, New York, NY.

GRANTS AND AWARDS

NSF Doctoral Dissertation Research Improvement Grant ($39,000)	2025–2027
UCLA Dissertation Year Award ($30,000)	2025–2026
UCLA Graduate Dean's Scholar ($14,500)	2021–2023

Introduction

The question of how conditions encountered during development become biologically embedded and propagate into adult phenotype sits at the intersection of life history theory, developmental biology, and evolutionary biology. In short-lived species with simple social structures, early environments provide cues about future conditions that organisms use to calibrate their physiology and behavior accordingly. The resulting phenotype either fits the anticipated environment or suffers a mismatch cost. But this logic strains under the specific conditions that characterize long-lived, socially complex species. When the juvenile period spans years and the adult lifespan spans decades, the correlation between the environment of development and the environment of reproduction may be prohibitively low. Moreover, the juvenile phenotype itself may be under direct selection, meaning that traits persisting into adulthood could reflect selection on juvenile performance rather than adaptive calibration for the adult environment (Stamps, 2003). When social structure is fluid, coalitionary, and cognitively demanding, behavioral flexibility may be a more reliable fitness strategy than fixed developmental programming. And when fitness consequences play out over a lifetime rather than a single breeding season, the buffering capacity of long life histories against early adversity may itself constitute a form of adaptation. Consequently, it is not clear whether frameworks developed in laboratory rodents and captive primates which link early adversity through physiological programming to adult behavior and life history actually hold in wild, long-lived primates experiencing naturally occurring sources of developmental challenge.

This dissertation investigates that question in white-faced capuchin monkeys (*Cebus imitator*) at Lomas Barbudal Biological Reserve in Guanacaste, Costa Rica, using more than three decades of individual-based longitudinal data from the Lomas Barbudal Monkey Project. The empirical work is organized around a four-stage causal sequence: early-life adversity (ELA) → hypothalamic–pituitary–adrenal (HPA) axis regulation → behavioral phenotypes → life-history outcomes. Each stage of this sequence corresponds to a theoretical literature with its own predictions, methodological requirements, and open empirical questions. Across four chapters, I evaluate each link in turn, beginning with the direct life-history consequences of ELA, characterizing the candidate behavioral phenotypes through which adversity might shape fitness, establishing the methodology and dimensionality of HPA axis measurement in this population,

and finally integrating these components to test the full causal chain. The answer that emerges suggests that frameworks developed in other systems require substantial revision before they can account for developmental programming in wild, long-lived primates with complex coalitionary societies.

The theoretical scaffolding for this work spans several interconnected literatures: life history theory and its account of developmental plasticity; the adaptive calibration model (ACM) and related informational frameworks linking early adversity to HPA axis function; the somatic-state and developmental constraint alternatives to informational calibration; the animal personality and behavioral syndrome literature; the neuroendocrinology of coping styles; and the growing body of observational work on ELA and fitness in wild primates. Here I review each of these literatures in turn, drawing out the specific predictions they generate for the capuchin system and flagging the assumptions that may not survive translation from laboratory to field.

I. Life History Theory and Developmental Plasticity

A. Life History Trade-offs and Energetic Allocation

Life history theory begins with an observation about constraint. Because metabolic energy is finite and time is limited, investment in one component of fitness, be it growth, maintenance, or current reproduction, necessarily reduces investment in others (Roff, 2001; Stearns, 1989). Natural selection acts on these trade-offs, shaping the timing and intensity of growth, reproductive maturation, reproductive effort, and senescence in response to the ecological conditions a population faces (Stearns, 2000). The result is the remarkable diversity of life history strategies observed across the animal kingdom. Small-bodied species exhibit rapid maturation and high fecundity under conditions of high extrinsic mortality risk; large-brained, socially complex species occupying more stable ecological niches prolong development, exhibit low fecundity, and have extended lifespans (Charnov & Berrigan, 1993; Harvey & Clutton-Brock, 1985).

In long-lived primates, the key features of the life history landscape are slow development, heavy parental investment, and lifetime reproductive success that depends primarily on adult survival rather than fecundity (Harvey & Clutton-Brock, 1985; Isler & Van Schaik, 2014; van Schaik & Isler, 2012). Reproductive function in primates is tightly regulated by energetic state. Ovarian function, in particular, is sensitive to the balance between energy intake and expenditure,

suppressed by periods of energetic deficit even in the absence of extreme nutritional deprivation (Ellison, 2003; Jasienska & Ellison, 2004). This energetic coupling means that individual reproductive trajectories are not fixed species constants but are sensitive to the conditions individuals actually experience across their development. Early-life environments may therefore set trajectories that propagate across the full length of the reproductive career.

B. Developmental Plasticity and Its Adaptive Interpretation

Phenotypic plasticity is defined as the capacity of a single genotype to produce different phenotypes in response to environmental conditions and represents a fundamental feature of development (West-Eberhard, 2003). A particularly consequential form occurs when environmental inputs experienced during development produce lasting changes in adult phenotype. Unlike contextual plasticity, which allows ongoing adjustment across the lifespan, so-called developmental plasticity generates phenotypic differences that persist well beyond the period of environmental exposure (Fawcett & Frankenhuis, 2015; West-Eberhard, 2003). The permanence of these effects raises the question: under what conditions should natural selection favor developmental sensitivity to early environments, and when should early experiences shape adult phenotypes in ways that enhance fitness rather than merely reflect developmental history? Throughout this dissertation, I use "development" and "developmental" to refer to early life, from conception to reproductive maturation, a period during which organisms are hypothesized to be especially sensitive to environmental inputs that can have lasting consequences for adult phenotype (Stamps, 2003; Fawcett & Frankenhuis, 2015)[1].

The framework of adaptive developmental plasticity (ADP) proposes that developmental responses to early environmental inputs can be shaped by selection to improve fitness under the conditions in which the resulting phenotype is expressed (Bateson et al., 2014; Nettle & Bateson, 2015). Two distinct logics fall under this umbrella. The first is informational: early-life cues are treated as forecasts of future environments. The second is state based: early environments shape adult phenotype through their cumulative effect on the organism's internal condition, independent of any predictive relationship between early and late environments. The first is contained in 'Informational models' which include predictive adaptive response (PAR) frameworks, and

[1] This focus on early-life developmental plasticity does not imply that later experiences are unimportant, only that the empirical questions motivating this dissertation concern the persistent effects of early conditions.

assume that early-life cues carry reliable statistical information about future environmental conditions which allow developing organisms to match their phenotype to the environment they are likely to encounter as adults (Gluckman et al., 2005; Nettle et al., 2013). Textbook examples of informational ADP come from short-lived species with well-defined environmental contrasts, the classic case being defensive morphology development in *Daphnia* in response to predator kairomones (Tollrian, 1995), where the correlation between developmental cue and adult selective environment is high and the lag between the two is short.

In long-lived, slow-developing species, however, the informational logic runs into difficulty. For cue-based plasticity to evolve, early environmental inputs must reliably forecast conditions encountered years or decades later, and that reliability requirement becomes increasingly implausible as lifespans lengthen and ecological conditions fluctuate across generations (Frankenhuis et al., 2019; Nettle et al., 2013). Theoretical work has shown that the informational value of early cues decays rapidly under realistic ecological variability, such that selection for predictive calibration is weak or absent in species in which the juvenile period spans a substantial fraction of their generation time (Fawcett & Frankenhuis, 2015, Frankenhuis & Panchanathan, 2011; Donaldson-Matasci et al., 2013). This observation has prompted sustained interest in alternative mechanisms that do not depend on cue reliability (Nettle & Bateson, 2015; Frankenhuis et al., 2019).

C. The Somatic State Alternative

An alternative framework holds that early-life environments shape adult phenotype through their effects on organismal state (Nettle et al., 2013; Nettle & Bateson, 2015). State-based models of ADP trace their lineage to state-dependent life history theory, which formalized the insight that optimal behavioral and reproductive decisions depend on an individual's current condition as well as on external circumstances (Clark, 1994; Houston & McNamara, 1999; McNamara & Houston, 1996). Within this tradition, "state" refers to any internal variable that *carries over across time and influences the optimal course of action*, which may include energy reserves, body size, dominance rank, immune status, accumulated physiological damage, or (as I argue here) the functional parameters of regulatory systems such as the hypothalamic-pituitary-adrenal (HPA) axis. Because state variables persist and shape decision-making across many contexts (Dingemanse & Wolf, 2010), modest developmental perturbations in state can generate

substantial and enduring phenotypic differences without any requirement that early environments predict later conditions (Luttbeg & Sih, 2010; Sih et al., 2015).

Under an informational account, developmental responses to early conditions are adaptive if and only if they calibrates the developing organism to an external environment that reliably materializes (Nettle et al., 2013). If the adult external environment differs from what early cues implied, the individual suffers a mismatch cost. By contrast, under a state-based account, the question of cue reliability does not arise because the developmental input does not function as a forecast of future conditions. Instead, early experience leaves cumulative traces that become embedded in the organism's internal environment – its somatic condition, including energetic reserves, immune function, and physiological set-points (Nettle et al., 2013; West-Eberhard, 2003). Subsequent behavior is then optimized conditional on that internal environment, not on any expected external state. Crucially, an individual entering adulthood with depleted energetic reserves, compromised immune function, or altered physiological set-points is not misreading the world outside; rather, its own body is the relevant milieu for decision-making, in the sense that downstream decisions must be made relative to the state it carries. This reframing implies that developmentally induced phenotypic shifts can be adaptive even when they lead to shortened lifespans or other outcomes that would appear maladaptive under a constraint interpretation (Nettle et al., 2013).

State-based ADP is best understood not as a rejection of adaptive interpretations of developmental plasticity, but as a reformulation of what qualifies a developmental response as adaptive. Where informational models require a correspondence between developmental cue and future environment, state-based models require only that behavioral and physiological decision-making be sensitive to the organism's own condition - a much weaker and more easily satisfied assumption for long-lived species. For the same reason, state-based ADP can be empirically difficult to distinguish from pure developmental constraint: in both cases, early experience produces lasting physiological effects that shape adult phenotype, and the difference lies in whether downstream responses to those effects are themselves under selection (Nettle & Bateson, 2015). This interpretive challenge runs through all four empirical chapters of this dissertation and motivates the strategy of evaluating each link in the hypothesized causal chain explicitly rather than inferring the full chain from any single observed outcome.

D. The Hypothalamic-Pituitary-Adrenal (HPA) Axis as a Candidate State Variable

1. The HPA axis

The HPA axis is the primary neuroendocrine system that mobilizes energy in response to challenges, regulates recovery from stress, and modulates a wide range of physiological and behavioral processes (Sapolsky et al., 2000). Briefly, it consists of a cascade of hormonal signals: hypothalamic release of corticotropin-releasing hormone (CRH) stimulates pituitary release of adrenocorticotropic hormone (ACTH), which in turn triggers adrenal release of glucocorticoids (cortisol in primates) into circulation. Glucocorticoids then act on target tissues throughout the body, including the brain, where they regulate metabolism, immune function, and behavior, and exert negative feedback on the hypothalamus and pituitary to terminate the stress response (Sapolsky et al., 2000; McEwen, 1998).

Several features of the HPA axis function together make it a defensible candidate for the role of somatic state variable in the theoretical framework developed above. First, individual differences in HPA axis activity are repeatable across time within individuals, satisfying the persistence requirement (see Schoenemann & Bonier, 2018; Taff et al., 2018; discussed further in section II.D.2 below). Second, glucocorticoids act pleiotropically across tissues and organ systems, regulating energy mobilization, immune function, growth, and reproductive physiology, and are therefore positioned to influence a wide range of fitness-relevant decisions rather than affecting only a single narrow outcome (Sapolsky et al., 2000). Third, decades of experimental and observational work have established that variation in early environments produces lasting changes in glucocorticoid regulation (Lupien et al., 2009; Meaney, 2001; Murphy et al., 2022; Speranza et al., 2024). Fourth, and especially important for state-based frameworks, persistent elevation of circulating glucocorticoids is associated with cumulative physiological costs through allostatic load, such that an individual's HPA axis profile indexes not only its current regulatory capacity but also accumulated physiological wear (McEwen, 1998; McEwen & Wingfield, 2003).

Glucocorticoid activity can be characterized along at least two temporal dimensions relevant to this dissertation. Tonic or baseline glucocorticoid concentrations reflect the ongoing physiological state and integrate current energetic condition, circadian rhythms, and individual differences in regulatory set-points (Sapolsky et al., 2000; Romero, 2004). Acute glucocorticoid responses capture the capacity to mobilize resources in response to immediate challenge, while

recovery from activation reflects the efficiency of negative feedback regulation (Wingfield et al., 1998; Zimmer et al., 2019). These dimensions are at least partially independent - an individual with elevated tonic glucocorticoids need not mount a strong acute response, and vice versa (Mentesana & Hau, 2022; Romero, 2004). This independence is central to the adaptive calibration framework discussed in Section III, which proposes that tonic and reactive HPA activity can be configured independently in response to developmental conditions.

In wild primates, glucocorticoids are commonly measured non-invasively in feces. Fecal glucocorticoid (fGC) metabolites integrate hormone production and excretion over several hours, making them well suited for characterizing tonic or baseline activity but presenting challenges for measuring acute reactivity, which requires knowledge of the species-specific lag between HPA activation and peak fGC appearance (Palme, 2019; Sheriff et al., 2011; Touma & Palme, 2005). For white-faced capuchins, empirical work has established a peak fGC excretion window of approximately 6–10 hours post-stressor (Beehner et al., 2022), a finding that is central to the rolling-window methodology developed in Chapter 3.

2. Repeatability of Glucocorticoid Measures

For HPA axis function to serve as a somatic state variable mediating between early experiences and adult life history, it must exhibit stable individual differences. If two measurements of an individual's glucocorticoid activity taken at different times are no more similar to one another than measurements from different individuals, there is no individual-level signal for developmental experience to have shaped and no stable trait through which adult decisions are conditioned. Stable individual differences are thus a conceptual prerequisite for treating HPA axis function as a mediating state.

Meta-analytic synthesis of repeatability estimates for glucocorticoid measures finds mean intraclass correlation coefficients (ICCs) in the range of approximately 0.20 to 0.45 for baseline or tonic measures, with lower and more variable estimates for stress-induced or reactive measures (Schoenemann & Bonier, 2018; Taff et al., 2018). These estimates are modest by the standards of morphological traits, but consistent enough across studies to support the claim that individual differences in glucocorticoid function are real and biologically meaningful. They align with the broader behavioral repeatability literature, in which mean repeatability estimates of 0.30 to 0.40 are typical for traits measured under realistic field conditions (Bell et al., 2009). Baseline

glucocorticoid measures are consistently more repeatable than stress-induced measures, plausibly because acute responses are more susceptible to context-specific factors that introduce additional within-individual variance (Schoenemann & Bonier, 2018; Taff et al., 2018). Fecal glucocorticoid measures yield repeatability estimates broadly comparable to those from blood and saliva, supporting their use as indicators of individual differences in HPA axis function under field conditions (Acker et al., 2018; Hau et al., 2022; Pritchard et al., 2023).

II. Early-Life Adversity: Frameworks, Measurement, and Evidence

A. Defining Early-Life Adversity

Early-life adversity (ELA), as used in this dissertation, encompasses events and conditions encountered during development that disrupt an organism's health, growth, physiological function, or social relationships in ways with lasting consequences. This broad definition is intentional as the empirical literature encompasses a wide range of developmental perturbations, from experimentally imposed manipulations in laboratory animals to naturally occurring ecological and social stressors in wild populations. Restricting the definition of ELA prematurely risks excluding forms of adversity whose effects may differ in theoretically informative ways. ELA is associated with a range of downstream effects on health, behavior, and life history outcomes across the lifespan (Lu et al., 2019; Shalev & Belsky, 2016; Sloboda et al., 2009) and has been shown to shorten longevity. However, these associations are not uniform - their direction, magnitude, and mechanism vary across adversity types, developmental windows, and species in ways that motivate targeted analysis rather than global summary (Brenhouse & Bath, 2019; Entringer & Heim, 2024; Lu et al., 2019; Waters & Gould, 2022).

B. Predictive Adaptive Response versus Developmental Constraint

The theoretical debate about how early adversity should affect life history trajectories has been organized primarily around two competing frameworks. Predictive adaptive response (PAR) models propose that ELA acts as a reliable informational cue: exposure to harsh or unpredictable developmental conditions signals elevated extrinsic mortality risk, and organisms adaptively respond by accelerating reproductive maturation and effort - pursuing a "live fast, die young" strategy that maximizes short-term reproductive output in anticipation of a shortened lifespan (Del Giudice, 2014; Ellis et al., 2022; Lea et al., 2017). From this perspective, adversity-exposed

individuals who reproduce earlier are implementing a conditional life history strategy appropriate to their developmentally forecast environment. Empirical support for reproductive acceleration following early-life adversity exists in several non-human systems. In invertebrates, *Daphnia* exposed to predator kairomones reproduce earlier at smaller body size and exhibit greater reproductive investment (Chakri et al., 2010; Stibor, 1992). In rodents, adverse early rearing conditions, such as low maternal licking and grooming, enhance early reproductive fitness, including increased sexual receptivity and precocious sexual maturity (Hengartner, 2017). In ground squirrels, offspring of mothers exposed to cues of high population density grew faster and established reproductive territories earlier, but only when adult conditions matched the developmental forecast (Lu et al., 2019; Sheriff et al., 2010). In captive macaques, exposure to maternal separation or harsh rearing has been linked to earlier age at first birth, consistent with predictive adaptive responses (Maestripieri, 2005). However, the majority of evidence from longer-lived vertebrates is more equivocal: experimental elevation of stress hormones in early life in starlings reduced longevity without accelerating reproduction (Monaghan & Haussmann, 2015), and tests of PAR predictions in long-lived mammals have more often supported developmental constraints than adaptive forecasting (Lu et al., 2019).

Constraint models offer a competing account. Rather than functioning as an adaptive cue, ELA imposes direct physiological limitation that may delay or suppress reproductive development regardless of any environmental forecast (Monaghan, 2008; Weibel et al., 2020). In primates, reproduction is governed by the energy available above the demands of maintenance and growth; when adverse early conditions deplete energetic state, reproductive function is downregulated through the hypothalamic-pituitary-gonadal axis until sufficient reserves accumulate (Ellis, 2004; Ellison, 2003; Hill & Elias, 2018). On this account, any association between early adversity and altered reproductive timing reflects the energetic and physiological costs of adversity rather than adaptive calibration. Contemporary energetic models of reproductive function emphasize that ovarian activity is sensitive to energy expenditure and availability even in the absence of extreme malnutrition (Ellison, 2003; Jasienska & Ellison, 2004; Vitzthum, 2008).

Evidence from long-lived primates has generally favored the constraint interpretation. In wild baboons, cumulative early adversity profoundly reduces adult lifespan but does not reliably accelerate reproduction; adversity appears to constrain rather than accelerate life history, pushing

individuals away from fitness-optimal trajectories (Tung et al., 2016; Weibel et al., 2020). In Cayo Santiago rhesus macaques, severe ecological adversity early in life was associated with delayed rather than accelerated reproductive debut (Luevano et al., 2022). These findings suggest that the energetic and physiological costs of adversity tend to dominate over any predictive calibration signal, challenging the generality of PAR models outside of short-lived systems. Chapter 1 of this dissertation tests these competing frameworks in wild female capuchins.

C. Dimensional Models of Adversity

A longstanding limitation of the ELA literature has been the widespread use of cumulative risk indices - scores that count adverse events while treating all forms of adversity as equivalent and additive (McLaughlin & Sheridan, 2016). This approach has practical advantages in large cohort studies but risks obscuring the distinct causal pathways through which different forms of adversity operate. Dimensional models of adversity have been developed to address this limitation by identifying the core features that distinguish adversity types and generate different developmental outcomes (Belsky et al., 2012; Ellis et al., 2022; Humphreys & Zeanah, 2015; McLaughlin & Sheridan, 2016).

The threat/deprivation framework, developed primarily in developmental neuroscience, distinguishes between adversity involving harm or threat of harm and adversity involving absence of expected stimulation, arguing that these dimensions engage different neurobiological mechanisms and produce divergent developmental outcomes (McLaughlin et al., 2021; McLaughlin & Sheridan, 2016; Sheridan & McLaughlin, 2014). A complementary evolutionary framework distinguishes harshness, defined as extrinsic sources of morbidity and mortality, from unpredictability, defined as stochastic temporal variability in those conditions, arguing that each dimension carries different implications for the reliability of developmental calibration (Brumbach et al., 2009; Ellis et al., 2022). This dissertation operationalizes adversity in ways that honor these distinctions, treating social instability, maternal loss, and ecological harshness as distinct adversity types whose effects on HPA axis function, behavior, and life history may operate through different mechanisms and vary across different timescales.

D. Sensitive Windows in Development

The effects of early-life adversity on HPA axis function and life history depend not just on whether adversity occurred but also on when that adversity occurred. The developing brain and neuroendocrine system exhibit successive periods of heightened sensitivity to environmental input, with different regulatory circuits undergoing rapid maturation at different ages and consequently being differentially vulnerable at different developmental points (Lupien et al., 2009). Sensitive windows represent the optimal temporal trade-off between the reliability of incoming environmental cues (which decreases as more time elapses between cue and adult environment) and the long-term fitness benefits of phenotypic specialization, which require early commitment to a developmental trajectory (English & Barreaux, 2020; Fawcett & Frankenhuis, 2015).

Empirical evidence for timing-specific effects has accumulated across rodent, captive primate, and human studies. Prenatal stress exposure produces lasting HPA axis alterations that differ in magnitude and pattern from those produced by postnatal manipulations (Coe et al., 2003; Weinstock, 2008), with similar timing effects documented in humans (Glover et al., 2018). Chronic, low-grade adversity and brief, intense stressors appear to produce different patterns of developmental recalibration, with chronic exposure more often producing blunted reactivity and acute severe exposure more often producing sensitization (Heim & Nemeroff, 1999; Lupien et al., 2009). Composite adversity measures computed across the full developmental period risk averaging across windows whose effects differ in both direction and magnitude. This dissertation therefore examines adversity effects separately across three developmental periods to characterize window-specific effects: the late prenatal, first year of life, and first five years of life.

E. Early-Life Adversity and Fitness in Wild Primates

A growing body of observational work in wild primate populations has documented the long-term consequences of naturally occurring ELA for adult phenotype and life history. The most robust finding is that cumulative early adversity predicts shortened adult lifespan: in wild baboons, females who experienced high early adversity died a median of ten years earlier than less-affected counterparts (Tung et al., 2016; Zipple et al., 2019), and effects on survival have been documented for specific adversity types including maternal loss, low maternal social status, drought exposure, and group instability across multiple primate species (Lea et al., 2015; Strauss et al., 2020). The effect sizes are comparable to those of sex on adult lifespan in many populations, establishing that

developmental sensitivity to early environmental conditions is a robust and consequential feature of primate biology.

The evidence linking naturally occurring ELA to HPA axis function in wild primates is considerably thinner. The broader wild primate stress physiology literature has been dominated by studies of current social rank and glucocorticoid levels rather than by studies of developmental plasticity (MacLeod et al., 2023; Sapolsky et al., 2000). The most frequently cited demonstrations of an adversity–HPA axis link in wild primates come from baboon work documenting associations between cumulative adversity indices and adult glucocorticoid concentrations (Rosenbaum et al., 2020), but the reported effects are modest in magnitude, inconsistent across populations, and in some cases detectable only for particular adversity subcomponents (Brown et al., 2024; Onyango et al., 2008; Patterson et al., 2021). Effects on acute HPA axis reactivity, as distinct from tonic levels, are even less well characterized, and the literature has not converged on a replicable pattern (Brown et al., 2024).

Several considerations help explain this asymmetry between the relatively robust lifespan literature and the weaker HPA literature. Cumulative adversity indices were developed to maximize predictive power for mortality outcomes and may not be equally suited to detecting effects on physiological intermediates. The interpretive chain from developmental exposure to adult hormone concentration is longer and more susceptible to measurement noise than the chain from exposure to survival, with intervening contextual variables introducing substantial within-individual variance in glucocorticoid measures. And sample size constraints are more severe for HPA axis studies than for lifespan studies, because hormone measurement requires sampling across the lifespan. These considerations mean that attempts to detect ELA effects on HPA axis function in wild primates must take seriously the possibility that such effects are modest but that they will require analysis at the level of specific adversity types and developmental windows, and furthermore that null findings require careful disambiguation from low-power non-detections.

III. The Adaptive Calibration Model

A. Theoretical Roots in Life History Theory

The Adaptive Calibration Model (ACM), introduced by Del Giudice, Ellis, and Shirtcliff (2011), offers the most developed theoretical account of how individual differences in stress

responsivity should be understood from an evolutionary developmental perspective. Its central proposal is that the stress response system is not primarily a regulatory mechanism whose variation reflects dysregulation or pathology, but an evolved system for calibrating an organism's physiology and behavior to the developmental conditions it experiences. The ACM's intellectual origins lie in life history theory and in earlier frameworks linking developmental experience to reproductive strategy (Belsky et al., 1991; Chisholm et al., 1993; Ellis, 2004). The ACM's distinctive contribution is to identify the HPA axis as the proximate physiological hub through which environmental conditions become biologically embedded and translated into life history strategies (Del Giudice et al., 2011; Ellis et al., 2017). This lineage positions the ACM squarely within the informational tradition of ADP, assuming that early-life cues carry reliable information about future conditions and that the developing stress response system is selected to integrate those cues into calibrated phenotypes - an assumption whose tenability in long-lived wild primates is a central concern of this dissertation.

B. Beyond the Allostatic Load Model

The ACM stands in contrast with the Allostatic Load Model (ALM), the dominant framework for understanding HPA axis function in biomedical research (McEwen, 1998; McEwen & Stellar, 1993). The ALM emphasizes cumulative physiological costs of repeated or chronic stress response activation: sustained glucocorticoid exposure compromises immune function, cardiovascular health, and metabolic regulation, and departures from a "healthy" HPA axis profile are typically interpreted as dysregulation with deleterious consequences. The ACM does not deny that prolonged HPA axis activation carries physiological costs but disputes the inference that all departures from the standard profile reflect dysregulation. Del Giudice and colleagues (2011) argue that some departures are better understood as adaptive recalibrations to environments in which different regulatory profiles carry different fitness consequences. The ALM and the ACM are most usefully understood as complementary: the ALM characterizes the physiological consequences of HPA axis activation, while the ACM proposes a functional interpretation of why developmental sensitivity has evolved. Distinguishing between calibration and damage requires evidence about whether the resulting phenotypes confer fitness advantages in the environments in which they are expressed - evidence that long-term observational studies of wild primates are, in principle, positioned to provide.

C. The Four-Pattern Taxonomy

The ACM's most distinctive empirical claim is that the joint distribution of tonic and reactive HPA axis activity is structured into a small number of qualitatively distinct profiles, each associated with particular developmental antecedents and strategic orientations. Del Giudice and colleagues (2011) describe four such profiles: Sensitive (moderate tonic activity, high responsivity; hypothesized to develop in safe and supportive environments); Buffered (low tonic activity, moderate responsivity; expected under intermediate conditions); Vigilant (elevated tonic activity, high responsivity; expected under moderate to high stress); and Unemotional (low or blunted reactivity despite chronic adversity; expected under severe and unpredictable stress and supporting behavioral disengagement). These profiles are theoretically defined by the joint distribution of tonic and reactive HPA axis parameters rather than by either alone, implying that empirical tests require methods capable of characterizing variation along both dimensions simultaneously.

Empirical support for the ACM in human samples is generally positive but qualified: studies using person-centered approaches have identified between three and five empirically derived HPA axis profiles whose features partially align with the model's predicted patterns (Del Giudice et al., 2012; Ellis et al., 2017; Peckins et al., 2015). Recent reviews have concluded that the model's predictions are most robustly tested as continuous associations between adversity exposure and configurations of tonic and reactive activity rather than as categorical group memberships (Ellis et al., 2017). Critically, virtually all empirical tests have been conducted in humans using laboratory-derived cortisol measures, and the model was developed within an informational ADP tradition whose assumptions have not been systematically evaluated in long-lived non-human species. Testing the ACM in wild capuchins therefore addresses a gap that the human-centered empirical literature has not yet filled.

IV. Developmental Programming of the HPA Axis

A. Maternal Care and Glucocorticoid Receptor Regulation

The empirical case for treating the HPA axis as a developmentally programmable system rests heavily on foundational experimental work in laboratory rodents demonstrating that early environmental inputs produce lasting changes in glucocorticoid regulation through identifiable

molecular mechanisms. This work represents the most mechanistically detailed example of a broader phenomenon (developmental programming of the HPA axis) that has now been documented across diverse vertebrate taxa (e.g., Banerjee et al., 2012; Crespi & Denver, 2005; Ensminger et al., 2018; Taborsky, 2016). Meaney and colleagues established that naturally occurring variation in maternal care behavior produces stable individual differences in HPA axis function that persist into adulthood (Liu et al., 1997; Meaney, 2001). Offspring of high-licking mothers exhibit lower tonic corticosterone levels, more rapid recovery from acute stressors, and reduced anxiety-like behavior compared to offspring of low-licking mothers; these differences are reliably reproduced when offspring are cross-fostered, demonstrating mediation by maternal behavior rather than genetic transmission (Francis et al., 1999; Liu et al., 1997). The mechanism has been well characterized. Variation in maternal licking and grooming alters methylation of the glucocorticoid receptor gene promoter in the hippocampus, with offspring of high-licking mothers showing reduced methylation, increased glucocorticoid receptor expression, and more efficient negative feedback regulation (Weaver et al., 2004; 2006). Subsequent work has extended these findings to prenatal stress, postnatal handling, and maternal separation, all of which produce lasting HPA axis alterations through partially overlapping epigenetic and neurochemical pathways (Champagne et al., 2009; Lupien et al., 2009; Meaney, 2001; Szyf et al., 2008).

B. Primate Models of Early-Life Adversity

The translation of developmental programming principles from rodents to primates has been pursued primarily through experimental and seminaturalistic work in macaques and marmosets. Studies of rhesus macaques subjected to maternal separation, peer-rearing, or unstable foraging demand paradigms have documented persistent alterations in HPA axis activity emerging in infancy and persisting into the juvenile and subadult periods (Coplan et al., 1996; Sanchez, 2006). Maternally separated or peer-reared infants exhibit elevated tonic cortisol levels, exaggerated cortisol responses to novel stressors, and altered diurnal cortisol rhythms relative to mother-reared controls (Buschdorf & Meaney, 2016; Meyer & Hamel, 2014; Parker & Maestripieri, 2011). Work in abused infant macaques has demonstrated that the specific character of early adversity, not merely its presence, shapes the resulting HPA axis profile (Maestripieri et al., 2006; McCormack et al., 2009). Studies in common marmosets have produced broadly convergent findings (Pryce et al., 2010). These findings from multiple primate taxa strengthen the

case that developmental programming of the HPA axis is a general feature of mammalian biology rather than a peculiarity of laboratory rodents.

However, the primate experimental literature has important limitations for predicting what should be observed in wild populations. The forms of adversity studied in laboratory contexts (maternal separation, peer rearing, abusive treatment) do not necessarily correspond to the range of developmental challenges that wild primates routinely face. Wild infants experience maternal loss, takeover-related infanticide risk, group instability, ecological perturbation, and chronic resource limitation, and whether these naturalistic forms of adversity produce HPA axis effects analogous to those documented in experimental settings cannot be determined from experimental data alone. The experimental literature provides strong proof of principle that early environments can program HPA axis function (Wood et al., 2004; McCormack et al., 2025; Sanchez, 2006; Meaney, 2001; Brown et al., 2024; Francis, 2002), but the translation of these principles to wild populations requires observational data from free-ranging animals experiencing naturally occurring adversity.

C. Timing, Duration, and Specificity of Developmental Effects on HPA axis Function

The consequences of early adversity for HPA axis function depend critically on when, for how long, and in what form the adversity occurs. Prenatal stress exposure produces lasting HPA axis alterations that differ in magnitude and pattern from those produced by postnatal manipulations (Coe et al., 2003; Weinstock, 2008), with similar timing effects documented in humans (Glover et al., 2018). Among post-natal exposures, chronic, low-grade adversity and brief, intense stressors appear to produce different patterns of HPA axis recalibration, with chronic exposure more often producing blunted reactivity and acute severe exposure more often producing sensitization (Heim & Nemeroff, 1999; Lupien et al., 2009). These timing- and duration-specific effects have direct methodological consequences: operationalizations of ELA that conflate developmentally distinct exposures risk averaging across effects that differ in direction or magnitude, potentially attenuating effects that would be detectable if each adversity type and developmental window were analyzed separately.

V. Animal Personality, Behavioral Syndromes, and the Boldness–Aggression Axis

A. Animal Personality and Repeatability

The study of animal personality emerged from the recognition that consistent individual differences in behavior are not transient noise around an adaptive optimum but are biologically meaningful targets of selection in their own right (Bell, 2007; Dingemanse & Réale, 2005; Réale et al., 2007; Sih et al., 2004; Wolf et al., 2007). Animal personality is formally defined as repeatable individual differences in behavior that remain consistent across time and contexts, with repeatability (R) (the proportion of total phenotypic variance attributable to stable among-individual differences) serving as the core empirical metric (Bell et al., 2009; Réale et al., 2007). A meta-analytic review of repeatability estimates in free-living animals found a mean R of approximately 0.37 across a broad range of behavioral domains including aggression, activity level, exploration, anti-predator behavior, foraging, courtship, and parental care, establishing a cross-taxa quantitative benchmark against which individual study estimates can be evaluated (Bell et al., 2009).

Several theoretical frameworks have been proposed to explain why consistent individual differences in behavior should persist despite natural selection acting to eliminate suboptimal fixed strategies. Life history trade-offs can favor the maintenance of among-individual behavioral variation when the costs and benefits of behavioral strategies differ across environments (Wolf et al., 2007; Wolf & Weissing, 2012). State-dependent feedback models propose that initial differences in condition can be amplified through positive feedback as individuals gain experience, generating persistent variation in behavioral type through developmental dynamics rather than requiring a direct genetic basis (Houston & McNamara, 1999; Luttbeg & Sih, 2010; J. A. Stamps, 2007). Pace-of-Life Syndrome (POLS) frameworks link behavioral tendencies to life history strategies, predicting that behavioral types should

covary with suites of physiological and demographic traits in ways that reflect the fundamental trade-off between current and future reproduction (Biro & Stamps, 2010; Réale et al., 2007). These frameworks share the insight that individual variation in behavior is not an epiphenomenon but may itself be a target of selection.

B. Behavioral Syndromes and the Boldness–Aggression Axis

When multiple behavioral traits covary predictably at the individual level such that an individual's behavioral type in one domain predicts its behavior in functionally distinct domains - the result is a behavioral syndrome (Sih, Bell, Johnson, et al., 2004). Among the syndromes documented across taxa, the boldness–aggression axis has attracted the most sustained empirical and theoretical attention (Bell, 2005; Sih, Bell, & Johnson, 2004). Boldness, defined as the propensity to take risks in potentially dangerous situations, and aggression, the initiation of agonistic conflict with conspecifics, are functionally distinct behaviors that are frequently correlated at the individual level, a pairing interpreted as evidence of shared proximate mechanism embedded in broader frameworks connecting behavioral strategies to life history trajectories (Réale et al., 2007, 2010).

The conditions under which the boldness–aggression syndrome emerges are not necessarily fixed species-level properties. Experimental work in sticklebacks has demonstrated that predation acts as an agent of correlational selection, selectively removing individuals whose boldness and aggression are mismatched, thereby tightening the boldness–aggression correlation among survivors (Bell, 2005; Bell & Sih, 2007). This ecological contingency implies that syndrome strength depends on the selective landscape in which a population is embedded, not on some fixed behavioral architecture inherent to the species. A critical further insight is that population-level correlations between boldness and aggression may conceal substantial

individual heterogeneity in how these traits are coupled (Dochtermann & Dingemanse, 2013). Some individuals may show strong positive coupling, others near-zero or even negative coupling, and a single population-mean correlation will be a poor summary of anyone's actual behavioral architecture. In species with complex social structures and variable life history contexts, where the costs and benefits of boldness and aggression shift substantially across rank, age, and reproductive state, such individual heterogeneity may itself be biologically meaningful.

C. Approaches to Measuring Personality in Primates

The dominant approach to studying individual differences in primate behavior has been adapted from human psychometrics, in which long-term observers rate individuals on personality items that are then subjected to factor analysis (Freeman & Gosling, 2010; Gosling, 2001). At Lomas Barbudal, this approach has identified five robust personality dimensions with high inter-rater reliability and meaningful associations with directly observed behavior and life history outcomes (Manson & Perry, 2013; Perry et al., 2017). However, composite factors derived from factor analysis can obscure the specific behavioral covariances that define syndromes. Applying a single label to phenotypically distinct behaviors (the *jingle* fallacy) and using different labels for behaviors that may reflect the same underlying process (the *jangle* fallacy) pose particular risks for syndrome research, and both have been documented in primate personality assessments (Block, 1995; Uher, 2011a, 2011b). In the Lomas Barbudal personality data, aggression loads across both Extraversion and Neuroticism rather than cleanly onto a single factor (Manson & Perry, 2013), meaning that even a strong behavioral coupling between risk-taking and aggressiveness might not be detectable in the factor structure. These considerations motivate a complementary bottom-up, trait-based approach that measures specific, ethologically valid

behaviors in defined functional contexts before testing their covariance (Carter et al., 2013; Hertel et al., 2020).

D. Developmental Influences on Behavioral Consistency

Early-life experience can shape both the consistency of individual behavioral traits and the degree to which traits are coupled across contexts (Edenbrow & Croft, 2012; Stamps & Groothuis, 2010). State-dependent feedback models predict that initial differences in developmental condition can be amplified through positive feedback as individuals gain experience, generating persistent variation in behavioral type even when the underlying developmental processes are similar across individuals (Houston & McNamara, 1999; Luttbeg & Sih, 2010; Stamps, 2007). The developmental canalization framework proposes more specifically that adversity during sensitive developmental periods may alter the degree of behavioral variance expressed by individuals, either reducing variance (canalization) as individuals become committed to a particular strategy, or increasing variance (reduced canalization) as adversity disrupts normative developmental trajectories (Royauté & Dochtermann, 2017; Stamps & Groothuis, 2010). The neuroendocrine stress response system, particularly the HPA axis, represents one pathway through which early experiences could generate lasting individual differences in both behavioral consistency and the strength of trait coupling (Sih, 2011). These developmental considerations connect the behavioral characterization pursued in Chapter 2 directly to the physiological analyses of Chapters 3 and 4.

VI. Neuroendocrine Correlates of Behavior and Life History

A. Coping Styles and the Proactive–Reactive Axis

The proactive–reactive coping styles model is the most widely invoked framework linking individual differences in HPA axis function to behavioral variation. Initially developed in

laboratory rodents and farmed species (Benus et al., 1991; Koolhaas et al., 1999), it has since been extended across a range of vertebrate taxa (Coppens et al., 2010; Koolhaas et al., 1999; Øverli et al., 2007). The model holds that individuals fall along an axis defined by coordinated physiological and behavioral responses to challenge. At the proactive end, individuals show low baseline glucocorticoid output, blunted acute HPA axis reactivity, and active behavioral engagement with stressors - including high rates of aggression, exploration, and routine formation (Ferreira et al., 2018; Øverli et al., 2007). At the reactive end, individuals display higher baseline glucocorticoids, more pronounced acute reactivity, and inhibited or withdrawn responses such as freezing, vigilance, and heightened behavioral flexibility in novel circumstances (Aerts, 2018; Briefer Freymond et al., 2015; Westrick et al., 2019). The coping styles framework provides the theoretical bridge that the ACM relies on to link HPA axis profiles to behavioral strategies and, through behavior, to life history outcomes. Extending this model to capuchins, developmentally calibrated HPA axis profiles are thus hypothesized to shape patterns of dispersal, competition, and reproductive investment through their behavioral expression.

B. Empirical Complications and the Limits of the Coping Styles Framework

Despite its influence, the coping styles framework has encountered substantial empirical complications when extended to wild populations and phylogenetically distant species. Meta-analytic synthesis of glucocorticoid–behavior associations in vertebrates finds that effect sizes are modest on average and frequently inconsistent across studies of the same species (Duckworth et al., 2023; Hau et al., 2016; Schoenle et al., 2021). The associations predicted by the coping styles framework - low tonic glucocorticoids associated with high aggression, blunted reactivity associated with bold exploration - are detectable in some study systems (Haller, 2014) but absent

or reversed in others (Pritchard & Palombit, 2022; Qu et al., 2018), and the conditions under which the predicted associations hold are not well characterized.

The assumption of a single proactive–reactive axis has also been criticized as overly rigid for cognitively flexible species. The standard coping styles model associates bold and aggressive phenotypes with behavioral rigidity and routine-bound strategies, and associates flexibility with reactive coping styles - a characterization that may capture the rodent data on which the framework was developed but does not generalize cleanly to species in which behavioral flexibility is itself a fitness-relevant trait (Coppens et al., 2010; Sih & Del Giudice, 2012). For a species like white-faced capuchins, whose ecological success depends on behavioral invention, extractive foraging requiring skill development, and socially learned traditions, boldness toward novel situations may facilitate rather than impede behavioral flexibility - a prediction that inverts the standard proactive–reactive framework (Fragaszy et al., 2004; Perry, 2011, 2020; Perry et al., 2003). Furthermore, behavioral trait correlations that define coping styles are not fixed across populations and ecological conditions (Sih & Del Giudice, 2012). For cognitively complex primates, the alignment between bold or aggressive behavior on one hand and behavioral rigidity on the other cannot be assumed but instead requires direct empirical evaluation.

C. Aggression Subtypes and Their Distinct Neuroendocrine Substrates

The behavioral category of "aggression" encompasses a broad range of phenomenologically and functionally distinct behaviors with partially distinct neuroendocrine substrates. Behavioral neuroendocrinology has increasingly emphasized that the endocrine correlates of aggression vary across social context, reproductive state, sex, species, and the specific aggressive behavior being measured, such that collapsing heterogeneous aggressive behaviors into a single construct can obscure rather than clarify their physiological bases

(Hirschenhauser & Oliveira, 2006; Nelson & Trainor, 2007; Rieger et al., 2022; Rosvall, 2013). The most influential distinction within this literature is between proactive (instrumental) and reactive (impulsive or defensive) aggression. Proactive aggression refers to low-arousal, goal-directed aggression deployed instrumentally in pursuit of resources, status, or other incentives, whereas reactive aggression refers to high-arousal, defensive, or retaliatory aggression elicited by provocation, frustration, or perceived threat (Dodge & Coie, 1987; Flanigan & Russo, 2019; Haller, 2018).

These two forms of aggression are associated with partially distinct profiles of HPA axis activity. Reactive aggression is more commonly associated with heightened physiological arousal and elevated acute HPA axis reactivity in response to provocation or social threat (Lopez-Duran et al., 2009; van Goozen et al., 2007; Van Honk et al., 2010). Proactive aggression, by contrast, has been associated with relatively blunted HPA axis function, including reduced basal cortisol levels and attenuated cortisol reactivity to challenge (Haller, 2018; Haller et al., 2004; Lopez-Duran et al., 2009). These distinctions generate specific empirical predictions for the capuchin data: if the coping styles–ACM framework applies, tonic HPA axis activity should predict proactive (unprovoked) aggression, while acute HPA axis reactivity should predict reactive (provoked) aggression. These predictions are tested explicitly in Chapter 4.

D. Boldness as a Behavioral Mediator

Boldness, broadly defined as the propensity to approach or persist in novel, risky, or potentially threatening situations rather than avoid them, is the second behavioral domain through which HPA axis function is hypothesized to influence life history outcomes (Réale et al., 2007; Sloan Wilson et al., 1994). In many species, boldness exhibits repeatability, ecological consequences, and links to fitness-relevant outcomes including foraging success, dispersal, and

survival (Bell et al., 2009; Réale et al., 2007; Smith & Blumstein, 2008; Wolf & Weissing, 2012). Within frameworks linking stress physiology to animal personality, boldness occupies a position parallel to but distinct from aggression: bold individuals are predicted to exhibit lower tonic glucocorticoid output and blunted acute reactivity, consistent with a proactive coping profile, whereas shy or cautious individuals are predicted to exhibit more reactive glucocorticoid profiles (Coppens et al., 2010; Koolhaas et al., 2007; Øverli et al., 2007).

Empirical support for this prediction is mixed. Studies across fish, birds, and mammals have sometimes found the predicted association between boldness and attenuated glucocorticoid activity (Schjolden et al., 2005; Seltmann et al., 2014; Thomson et al., 2011), but other studies report null or context-dependent relationships in which behavioral coping style and glucocorticoid activity do not covary (Alfonso et al., 2019; Clary et al., 2014; Westrick et al., 2019). The relationship between boldness and other behavioral traits, including aggression and behavioral flexibility, also varies across species, developmental histories, and ecological contexts rather than mapping uniformly onto a single proactive–reactive axis. For white-faced capuchins, where the male reproductive career may place contradictory demands on boldness and aggression across life history stages (Clark, 1994; Stamps, 2007), and where unprovoked aggression toward females may undermine the coalitionary alliances on which alpha tenure depends (Fedigan et al., 2021; Kajokaite et al., 2019; Perry, 1997, 1998), the coupling between boldness and aggression is likely to be complex and contingent rather than simple and fixed. Direct evaluation of HPA axis–boldness associations in this population, alongside the characterization of syndrome structure in Chapter 2, is therefore necessary before the behavioral pathway to life history outcomes can be evaluated.

E. From Behavioral Phenotypes to Life History Outcomes

The final link in the hypothesized pathway connecting HPA axis function to life history outcomes runs through behavior to dispersal, dominance acquisition, and reproductive success. In male capuchins, the relevant life history transitions are the timing of natal dispersal, the acquisition of alpha status in a new social group, and the duration of alpha tenure once acquired (Perry, 1998, 2012, 2017). Prior work on this population indicates that these transitions are linked to individual differences in behavioral phenotype: personality ratings have been associated with both the timing of male natal emigration and the speed with which males attain alpha status after dispersal (Perry et al., 2017). Alpha acquisition and tenure unfold in a social landscape characterized by male–male competition, coalitionary aggression, and relationship-dependent ally recruitment, making them plausible targets for behavioral mediation (Gros-Louis et al., 2003; Kajokaite et al., 2019; Perry, 2012; Perry et al., 2017). Proactive aggression and boldness therefore emerge as candidate behavioral mediators of male life history outcomes - the mechanistic links between developmental programming, physiological state, and fitness that Chapter 4 is designed to evaluate.

VII. Study System: White-Faced Capuchins at Lomas Barbudal

A. The Lomas Barbudal Monkey Project

This dissertation draws on data collected by the Lomas Barbudal Monkey Project (LBMP), a long-term field study of white-faced capuchin monkeys (*Cebus imitator*) established by Susan Perry in 1990. The study population inhabits a tropical dry forest in and around the Lomas Barbudal Biological Reserve, Guanacaste Province, Costa Rica (10°30'24" N, 85°22'17" W). Over more than three decades, the LBMP has accumulated individual-based longitudinal data on individually identified animals (including birth and death records, dominance history, reproductive history, group membership, behavioral observations, and fecal glucocorticoid

samples) at a depth and temporal resolution that makes it one of the most comprehensive long-term primate datasets available. The breadth and duration of this dataset are what make it possible to evaluate the full ELA → physiology → behavior → life history causal chain empirically, testing each link in the sequence across individuals whose developmental histories are known prospectively[2].

The physical environment of Lomas Barbudal is characterized by pronounced seasonality with a dry season lasting approximately five months from mid-December through mid-May, followed by a rainy season with total annual rainfall ranging from 500 to 2,500 mm, exhibiting high interannual variability tightly linked to large-scale climatic patterns including the El Niño Southern Oscillation. Phenological cycles at the site are tightly linked to seasonal rainfall. Canopy flushing is most prominent early in the rainy season, while fruiting tends to peak at the dry season's end and into the early wet months, a pattern consistent with observations from other Neotropical dry forests (Detto et al., 2018). This seasonality drives a cyclical pattern of resource availability in which the mid-to-late dry season represents a period of recurrent ecological and energetic stress for omnivorous foragers (Carnegie et al., 2011). Climatic perturbations - El Niño droughts or even exceptionally wet periods - can have acute and measurable impacts on nutritional intake, making rainfall deviation a biologically meaningful proxy for ecological adversity that appears repeatedly in this dissertation's analyses.

B. White-Faced Capuchin Biology

[2] Other landmark individual-based longitudinal studies, such as Santa Rosa Capuchin Project (since 1983; Fedigan & Jack, 2011), the Kibale Chimpanzee Project (since 1987; Watts, 2011), the Amboseli Baboon Research Project (since 1971; Alberts & Altmann, 2011), the Isle of Rum red deer project (since 1972; Clutton-Brock et al., 1982; Pemberton et al., 2022) and the Masai Mara spotted hyena project (since the 1980s; Rojas et al., 2022), have similarly accumulated decades of behavioral and demographic data, but none have concurrently measured, or evaluated the integration of, early-life adversity, adult HPA function, behavioral phenotypes, and life history outcomes in the same individuals.

White-faced capuchins (*Cebus imitator*) are among the most cognitively complex of the New World monkeys and present a combination of life history features that make them an informative study system for testing frameworks developed in humans and great apes. They are relatively long-lived for their body size with documented lifespans in wild populations exceed 30 years (Perry, 2012), and captive individuals have lived up to 55 years (Hakeem et al., 1996). They develop slowly, with females producing their first infant at approximately six to seven years of age and inter-birth intervals of approximately two to two and a half years (Hogan et al., 2019). They have large brains for their body size, with encephalization quotients comparable to those of great apes (Hartwig et al., 2011). They engage in sustained coalitionary politics involving alliance maintenance and cooperative territorial defense (Gros-Louis et al., 2003; Kajokaite et al., 2019; Perry, 1998, 2012). They rely on extractive foraging techniques that require skill development across an extended juvenile period (Perry, 2011). And they exhibit documented social learning and behavioral traditions that propagate across multi-decadal timescales (Perry, 2020; Perry et al., 2003). Taken together, these features make capuchins, like humans and great apes, a species in which the calibration logic of developmental programming faces serious challenges from the cue reliability problem, and in which behavioral flexibility and coalitionary social complexity introduce additional complications for frameworks that assume a rigid coupling of physiology to behavior.

Capuchin social groups are multi-male, multi-female with female philopatry and male dispersal. Reproductive success is highly skewed among males: alpha males sire the vast majority of offspring within a group, and achieving and maintaining alpha status is therefore the primary determinant of male fitness (Perry, 2012). Alpha male replacement events - "takeovers" - are often violent, competitive, and are frequently followed by infanticide against the offspring

of the deposed male, making them a major source of social instability for group members (Fedigan et al., 2021; Jack & Fedigan, 2004). Group fission events, in which a group permanently splits along matrilineal lines, represent a further source of social disruption. These naturally occurring events - takeovers, fissions, maternal loss, and seasonal resource fluctuation - constitute the forms of early-life adversity examined in this dissertation, and they span the range from direct social threat (prenatal exposure to takeovers) to energetic deprivation (rainfall deficit) to attachment disruption (maternal loss).

B. Capuchins as a Convergent System for Testing Developmental Frameworks

The empirical literature evaluating the ACM and related life history frameworks has been concentrated almost entirely in human populations, with supplementary work in laboratory rodents and captive or seminaturalistic macaques. Wild primate research can in principle address several key limitations of this literature. In long-term observational field studies, early-life environments are observed and recorded prospectively rather than reconstructed retrospectively from adult recall. Glucocorticoid sampling occurs under naturalistic conditions through non-invasive fecal methods that do not themselves activate the axis being measured, an important challenge in HPA axis research (Barnard & Sandusky, 2004; Langkilde & Shine, 2006; Romero, 2004). Life history outcomes are measured directly through longitudinal demographic data rather than self-reported or mediated by cultural confounds. And the stressors used to assess acute HPA axis reactivity - intergroup encounters, predator interactions, dominance challenges - are events with genuine fitness consequences that the stress response system can plausibly be argued to have evolved to address.

Within this comparative logic, platyrrhine primates such as white-faced capuchins occupy a particularly informative position. The HPA axis in anthropoid primates is homologous

to that of humans at a level of detail that supports close comparison of measurements and effect sizes. But the 40-million-year divergence between platyrrhines and catarrhines means that similarities between capuchin findings and human findings are more likely to reflect convergent outcomes of the developmental systems the framework proposes - rather than shared ancestry within the hominid or catarrhine lineage - providing a complementary line of evidence that cross-validates inference across primate taxa. The independent evolution of extended development, large-brained cognitive complexity, coalitionary politics, and extractive foraging in *Cebus imitator* thus makes this species both a challenging test case for developmental frameworks and, if those frameworks hold, a source of especially compelling evidence for their generality.

Non-invasive fecal sampling for glucocorticoid metabolites has been validated and routinely employed in this study population (Beehner et al., 2022). Fecal glucocorticoid (fGC) measurements differ from blood or saliva in that they integrate hormone metabolites produced and excreted over several hours, providing a temporally aggregated measure of HPA axis activity rather than an instantaneous snapshot (Palme, 2019; Sheriff et al., 2011). This integration is an asset for characterizing tonic or baseline HPA axis activity, smoothing over rapid pulsatile fluctuations, but introduces challenges for measuring acute stress reactivity, where accurate characterization requires knowledge of the species-specific lag between HPA axis activation and peak fGC appearance in feces (Touma & Palme, 2005). For white-faced capuchins, empirical work by collaborators has established a peak fGC excretion window of approximately 6-10 hours post-stressor - substantially shorter than the 12–48-hour range typical of larger primates and consistent with capuchins' rapid gut transit and smaller body size. This species-specific window is central to the rolling-window methodology developed in Chapter 3.

VIII. Sex-Specific Life History Pathways

This dissertation treats male and female life history outcomes separately. Chapter 1 examines life history outcomes in females, and Chapter 4 examines life history outcomes in males, while Chapters 2 and 3 characterize the behavioral and physiological phenotypes that serve as candidate mechanisms for both. This division reflects a theoretical claim about the pathways through which ELA translates into fitness consequences in each sex - pathways that differ in their proximate nature because of the way male and female reproductive success is achieved in this species.

In female capuchins, reproductive output is constrained primarily by the female's own physiology. Females invest heavily in each offspring through long gestation, extended lactation, and sustained parental care - the female herself bears the energetic and somatic costs of reproduction almost entirely. Consequently, variation in female fitness tracks variation in the female's physiological state - her energetic condition, somatic investment capacity, reproductive lifespan, and the hormonal regulation of reproductive function through the hypothalamic-pituitary-gonadal axis. ELA effects on female fitness may thus operate through physiological channels: adversity that depletes energetic reserves, impairs somatic condition, or alters the regulatory parameters of the HPA or HPG axes should translate into measurable consequences for survival and reproductive scheduling.

In male capuchins, the reproductive landscape is fundamentally different. Male reproductive success in this species is almost entirely determined by access to mating opportunities, and those opportunities are mediated through dominance rank and, especially, alpha status. Because alpha males sire the vast majority of offspring within a group (Perry, 2012), the fitness consequences of male behavioral phenotypes are concentrated in the competitive transitions to and from alpha status - natal dispersal, inter-group competition, alpha

attainment, and alpha tenure maintenance. The pathway from ELA to male fitness therefore runs not primarily through physiological regulation of energetic or reproductive processes, but through behavioral phenotypes that determine competitive success. A male's boldness in approaching novel or risky situations, his propensity for proactive versus reactive aggression toward competitors, and his ability to build and maintain the coalitionary alliances on which alpha tenure depends are the proximate determinants of his reproductive outcome. ELA matters for male fitness insofar as it shapes these behavioral traits - which is why the relevant male life history metrics are the behavioral transitions to and through alpha status rather than physiological reproductive scheduling.

This sex-by-pathway logic structures the dissertation in a principled way. Chapter 1 examines whether ELA predicts female mortality risk and the timing of reproductive milestones, the outcomes shaped primarily by female physiology. Because the proximate physiological mediators of female reproductive function (e.g., energetic condition, ovarian function) were not directly measured in this study, Chapter 1 evaluates the female pathway at the level of life history outcomes rather than testing the mediating mechanisms directly. Chapter 4 examines whether ELA programs HPA axis function, whether HPA axis function shapes behavioral phenotypes, and whether those behavioral phenotypes predict the competitive life history transitions that determine male fitness.

IX. Overview of the Dissertation

Chapter 1 tests whether early-life adversity predicts female mortality risk and the timing of reproductive milestones in wild white-faced capuchins, using longitudinal life history data and adversity measures across three developmental windows - the prenatal period, the first year of life, and the first five years of life. The chapter evaluates the competing predictions of predictive

adaptive response models and developmental constraint models with respect to reproductive timing and examines whether specific adversity types and developmental windows predict distinct effects on survival and reproduction. The analysis is restricted to females.

Chapter 2 characterizes individual differences in boldness and aggression as behavioral phenotypes using directly observed behavioral measures collected across the lifespan. The chapter evaluates whether boldness and aggression constitute repeatable individual traits in this population, whether they covary at the population level to form a boldness–aggression syndrome, and whether population-level syndrome absence or weakness conceals meaningful individual heterogeneity in how the two traits are coupled. It also evaluates the convergent validity of observer-assigned personality ratings with the directly measured behavioral traits. The results of this characterization serve as the necessary behavioral foundation for Chapter 4's tests of the behavioral mediation pathway from ELA to male fitness.

Chapter 3 addresses a fundamental methodological challenge in field endocrinology: how to extract biologically meaningful tonic and reactive HPA axis phenotypes from fecal glucocorticoid data collected opportunistically across many years of field study. The chapter develops and validates a rolling-window approach that leverages the known species-specific peak fGC excretion window to separate samples reflecting baseline HPA axis activity from those reflecting HPA axis activation in response to identifiable stressors (intergroup encounters). It characterizes the repeatability of each HPA axis dimension - establishing whether they meet the stability requirement for somatic state variables - and tests whether tonic and reactive HPA axis activity are empirically independent or correlated. The methodological framework and individual level HPA axis phenotypes produced in this chapter are the direct inputs to Chapter 4.

Chapter 4 integrates the foregoing components to evaluate the full hypothesized causal chain: ELA → HPA axis function → behavioral phenotypes → male life history outcomes. Using individual males with longitudinal glucocorticoid, behavioral, and life history data, the chapter tests each link in the chain explicitly: whether naturally occurring ELA programs adult HPA axis function, whether HPA axis phenotypes predict boldness and aggression, whether behavioral phenotypes predict male life history transitions including natal dispersal timing and the attainment and duration of alpha status, and whether any mediated pathway from ELA through physiology and behavior to fitness is supported. The chapter also evaluates direct effects of ELA on behavior and life history that bypass HPA axis function, allowing a direct comparison of mediated physiological pathways versus direct constraint or canalization effects. Together, the four empirical chapters address the overarching question of this dissertation: do the frameworks developed in laboratory and human research - linking early adversity through physiological calibration to adult behavioral phenotypes and fitness outcomes - hold when applied to wild, long-lived *Cebus imitator* at Lomas Barbudal? In answering this question, this dissertation moves beyond testing specific models to provide fundamental insight into how early-life conditions shape the trajectories of individual lives in a wild primate population. Understanding when and how adversity becomes biologically embedded (and when it does not) is essential not only for refining theoretical frameworks but also for understanding the ecological and evolutionary forces that shape variation in health, behavior, and fitness in natural populations.

References

Acker, M., Mastromonaco, G., & Schulte-Hostedde, A. I. (2018). The effects of body region, season and external arsenic application on hair cortisol concentration. *Conservation Physiology*, *6*(1). https://doi.org/10.1093/conphys/coy037

Aerts, J. (2018). Quantification of a Glucocorticoid Profile in Non-pooled Samples Is Pivotal in Stress Research Across Vertebrates. *Frontiers in Endocrinology*, *9*. https://www.frontiersin.org/journals/endocrinology/articles/10.3389/fendo.2018.00635/full

Alberts, S. C., & Altmann, J. (2011). The Amboseli Baboon Research Project: 40 years of continuity and change. In Long-term field studies of primates (pp. 261-287). Berlin, Heidelberg: Springer Berlin Heidelberg.

Alfonso, S., Sadoul, B., Gesto, M., Joassard, L., Chatain, B., Geffroy, B., & Bégout, M.-L. (2019). Coping styles in European sea bass: The link between boldness, stress response and neurogenesis. *Physiology & Behavior*, *207*, 76–85. https://doi.org/10.1016/j.physbeh.2019.04.020

Banerjee, S. B., Arterbery, A. S., Fergus, D. J., & Adkins-Regan, E. (2012). Deprivation of maternal care has long-lasting consequences for the hypothalamic–pituitary–adrenal axis of zebra finches. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 759-766.

Barnard, N. D. and Sandusky, C. (2004). Laboratory Routines Cause Animal Stress. Contemporary Topics in Laboratory Animal Science. 43(6), 42–51.

Bateson, P., Gluckman, P., & Hanson, M. (2014). The biology of developmental plasticity and the Predictive Adaptive Response hypothesis. *The Journal of Physiology*, *592*(11), 2357–2368. https://doi.org/10.1113/jphysiol.2014.271460

Beehner, J. C., Alfaro, J., Allen, C., Benítez, M. E., Bergman, T. J., Buehler, M. S., Carrera, S. C., Chester, E. M., Deschner, T., Fuentes, A., Gault, C. M., Godoy, I., Jack, K. M., Kim, J. D., Kolinski, L., Kulick, N. K., Losch, T., Ordoñez, J. C., Perry, S. E., … Wasserman, M. D. (2022). Using an on-site laboratory for fecal steroid analysis in wild white-faced capuchins. *General and Comparative Endocrinology*, *329*, 114109. https://doi.org/10.1016/j.ygcen.2022.114109

Bell, A. M. (2005). Behavioural differences between individuals and two populations of stickleback (*Gasterosteus aculeatus*). *Journal of Evolutionary Biology*, *18*(2), 464–473. https://doi.org/10.1111/j.1420-9101.2004.00817.x

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

Bell, A. M., & Sih, A. (2007). Exposure to predation generates personality in threespined sticklebacks (Gasterosteus aculeatus). *Ecology Letters*, *10*(9), 828–834. https://doi.org/10.1111/j.1461-0248.2007.01081.x

Belsky, J., Schlomer, G. L., & Ellis, B. J. (2012). Beyond cumulative risk: Distinguishing harshness and unpredictability as determinants of parenting and early life history strategy. *Developmental Psychology*, *48*(3), 662–673. https://doi.org/10.1037/a0024454

Belsky, J., Steinberg, L., & Draper, P. (1991). Childhood Experience, Interpersonal Development, and Reproductive Strategy: An Evolutionary Theory of Socialization. *Child Development*, *62*(4), 647–670. https://doi.org/10.1111/j.1467-8624.1991.tb01558.x

Benus, R. F., Bohus, B., Koolhaas, J. M., & van Oortmerssen, G. A. (1991). Heritable variation for aggression as a reflection of individual coping strategies. *Experientia*, *47*(10), 1008–1019.

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*(2), 187–215. https://doi.org/10.1037/0033-2909.117.2.187

Brenhouse, H. C., & Bath, K. G. (2019). Bundling the haystack to find the needle: Challenges and opportunities in modeling risk and resilience following early life stress. *Frontiers in Neuroendocrinology*, *54*, 100768. https://doi.org/10.1016/j.yfrne.2019.100768

Briefer Freymond, S., Bardou, D., Briefer, E. F., Bruckmaier, R., Fouché, N., Fleury, J., Maigrot, A.-L., Ramseyer, A., Zuberbühler, K., & Bachmann, I. (2015). The physiological consequences of crib-biting in horses in response to an ACTH challenge test. *Physiology & Behavior*, *151*, 121–128. https://doi.org/10.1016/j.physbeh.2015.07.015

Brown, E. R., Gettler, L. T., & Rosenbaum, S. (2024). Effects of social environments on male primate HPG and HPA axis developmental programming. *Developmental Psychobiology*, *66*(5), e22491. https://doi.org/10.1002/dev.22491

Brumbach, B. H., Figueredo, A. J., & Ellis, B. J. (2009). Effects of Harsh and Unpredictable Environments in Adolescence on Development of Life History Strategies. *Human Nature*, *20*(1), 25–51. https://doi.org/10.1007/s12110-009-9059-3

Buschdorf, J. P., & Meaney, M. J. (2016). Epigenetics/Programming in the HPA Axis. *Comprehensive Physiology*, *6*(1), 87–110. https://doi.org/10.1002/j.2040-4603.2016.tb00667.x

Carnegie, S. D., Fedigan, L. M., & Ziegler, T. E. (2011). Social and environmental factors affecting fecal glucocorticoids in wild, female white-faced capuchins (Cebus capucinus). *American Journal of Primatology*, *73*(9), 861–869. https://doi.org/10.1002/ajp.20954

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: What are behavioural ecologists measuring? *Biological Reviews*, *88*(2), 465–475. https://doi.org/10.1111/brv.12007

Chakri, K., Touati, L., Alfarhan, A. H., Al-Rasheid, K. A. S., & Samraoui, B. (2010). Effect of vertebrate and invertebrate kairomones on the life history of Daphnia magna Straus (Crustacea: Branchiopoda). *Comptes Rendus. Biologies*, *333*(11–12), 836–840. https://doi.org/10.1016/j.crvi.2010.09.004

Champagne, D. L., Ronald de Kloet, E., & Joëls, M. (2009). Fundamental aspects of the impact of glucocorticoids on the (immature) brain. *Seminars in Fetal and Neonatal Medicine, Uses and Abuses of Corticosteroids*, *14*(3), 136–142. https://doi.org/10.1016/j.siny.2008.11.006

Charnov, E. L., & Berrigan, D. (1993). Why do female primates have such long lifespans and so few babies? *Or* Life in the slow lane. *Evolutionary Anthropology: Issues, News, and Reviews*, *1*(6), 191–194. https://doi.org/10.1002/evan.1360010604

Chisholm, J. S., Ellison, P. T., Evans, J., Lee, P. C., Lieberman, L. S., Pavlik, Z., Ryan, A. S., Salter, E. M., Stini, W. A., & Worthman, C. M. (1993). Death, Hope, and Sex: Life-

History Theory and the Development of Reproductive Strategies [and Comments and Reply]. *Current Anthropology*, *34*(1), 1–24.

Clark, C. W. (1994). Antipredator behavior and the asset-protection principle. *Behavioral Ecology*, *5*(2), 159–170. https://doi.org/10.1093/beheco/5.2.159

Clary, D., Skyner, L. J., Ryan, C. P., Gardiner, L. E., Anderson, W. G., & Hare, J. F. (2014). Shyness–Boldness, but not Exploration, Predicts Glucocorticoid Stress Response in Richardson's Ground Squirrels (Urocitellus richardsonii). *Ethology*, *120*(11), 1101–1109. https://doi.org/10.1111/eth.12283

Clutton-Brock, T. H., Guinness, F. E., & Albon, S. D. (1982). *Red deer: behavior and ecology of two sexes*. University of Chicago press.

Coe, C. L., Kramer, M., Czéh, B., Gould, E., Reeves, A. J., Kirschbaum, C., & Fuchs, E. (2003). Prenatal stress diminishes neurogenesis in the dentate gyrus of juvenile Rhesus monkeys. *Biological Psychiatry*, *54*(10), 1025–1034. https://doi.org/10.1016/S0006-3223(03)00698-X

Coplan, J. D., Andrews, M. W., Rosenblum, L. A., Owens, M. J., Friedman, S., Gorman, J. M., & Nemeroff, C. B. (1996). Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult nonhuman primates exposed to early-life stressors: Implications for the pathophysiology of mood and anxiety disorders. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(4), 1619–1623.

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 B: Biological Sciences*, *365*(1560), 4021–4028. https://doi.org/10.1098/rstb.2010.0217

Crespi, E. J., & Denver, R. J. (2005). Ancient origins of human developmental plasticity. *American Journal of Human Biology*, *17*(1), 44-54.

Del Giudice, M. (2014). Early stress and human behavioral development: Emerging evolutionary perspectives. *Journal of Developmental Origins of Health and Disease*, *5*(4), 270–280. https://doi.org/10.1017/S2040174414000257

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The Adaptive Calibration Model of stress responsivity. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1562–1592. https://doi.org/10.1016/j.neubiorev.2010.11.007

Del Giudice, M., Hinnant, J. B., Ellis, B. J., & El-Sheikh, M. (2012). Adaptive Patterns of Stress Responsivity: A Preliminary Investigation. *Developmental Psychology*, *48*(3), 775–790. https://doi.org/10.1037/a0026519

Detto, M., Wright, S. J., Calderón, O., & Muller-Landau, H. C. (2018). Resource acquisition and reproductive strategies of tropical forest in response to the El Niño–Southern Oscillation. *Nature Communications*, *9*(1), 913. https://doi.org/10.1038/s41467-018-03306-9

Dingemanse, N. J., & Wolf, M. (2010). Recent models for adaptive personality differences: A review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 3947–3958. https://doi.org/10.1098/rstb.2010.0221

Dochtermann, N. A., & Dingemanse, N. J. (2013). Behavioral syndromes as evolutionary constraints. *Behavioral Ecology*, *24*(4), 806–811. https://doi.org/10.1093/beheco/art002

Dodge, K. A., & Coie, J. D. (1987). Social-information-processing factors in reactive and proactive aggression in children's peer groups. *Journal of Personality and Social Psychology*, *53*(6), 1146–1158. https://doi.org/10.1037/0022-3514.53.6.1146

Donaldson-Matasci, M. C., Bergstrom, C. T., & Lachmann, M. (2013). When unreliable cues are good enough. *The American Naturalist*, *182*(3), 313-327.

Duckworth, R. A., Chenard, K. C., Meza, L., & Beiriz, M. C. (2023). Coping styles vary with species' sociality and life history: A systematic review and meta-regression analysis. *Neuroscience & Biobehavioral Reviews*, *151*, 105241. https://doi.org/10.1016/j.neubiorev.2023.105241

Edenbrow, M., & Croft, D. P. (2012). Sequential hermaphroditism and personality in a clonal vertebrate: The mangrove killifish. *Behavioural Processes*, *90*(2), 229–237. https://doi.org/10.1016/j.beproc.2012.02.001

Ellis, B. J. (2004). Timing of Pubertal Maturation in Girls: An Integrated Life History Approach. *Psychological Bulletin*, *130*(6), 920–958. https://doi.org/10.1037/0033-2909.130.6.920

Ellis, B. J., Giudice, M. D., & Shirtcliff, E. A. (2017). The Adaptive Calibration Model of Stress Responsivity. In *Child and Adolescent Psychopathology, Third Edition* (pp. 237–276). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781394258932.ch8

Ellis, B. J., Sheridan, M. A., Belsky, J., & McLaughlin, K. A. (2022). Why and how does early adversity influence development? Toward an integrated model of dimensions of environmental experience. *Development and Psychopathology*, *34*(2), 447–471. https://doi.org/10.1017/S0954579421001838

Ellison, P. T. (2003). Energetics and reproductive effort. *American Journal of Human Biology*, *15*(3), 342–351. https://doi.org/10.1002/ajhb.10152

English, S., & Barreaux, A. M. (2020). The evolution of sensitive periods in development: Insights from insects. *Current Opinion in Behavioral Sciences, Sensitive and Critical Periods*, *36*, 71–78. https://doi.org/10.1016/j.cobeha.2020.07.009

Ensminger, D. C., Langkilde, T., Owen, D. A., MacLeod, K. J., & Sheriff, M. J. (2018). Maternal stress alters the phenotype of the mother, her eggs and her offspring in a wild-caught lizard. *Journal of Animal Ecology*, *87*(6), 1685-1697.

Entringer, S., & Heim, C. (2024). A Brief Historic Review of Research on Early Life Stress and Inflammation across the Lifespan. *Neuroimmunomodulation*, *32*(1), 24–35. https://doi.org/10.1159/000542676

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*(Suppl 1).

Fedigan, L. M., & Jack, K. M. (2011). Tracking neotropical monkeys in Santa Rosa: lessons from a regenerating Costa Rican dry forest. In *Long-term field studies of primates* (pp. 165-184). Berlin, Heidelberg: Springer Berlin Heidelberg.

Fedigan, L. M., Hogan, J. D., Campos, F. A., Kalbitzer, U., & Jack, K. M. (2021). Costs of male infanticide for female capuchins: When does an adaptive male reproductive strategy become costly for females and detrimental to population viability? *American Journal of Physical Anthropology*, *176*(3), 349–360. https://doi.org/10.1002/ajpa.24354

Flanigan, M. E., & Russo, S. J. (2019). Recent advances in the study of aggression. *Neuropsychopharmacology*, *44*(2), 241–244. https://doi.org/10.1038/s41386-018-0226-2

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The Complete Capuchin: The Biology of the Genus Cebus*. Cambridge University Press.

Francis, D., Diorio, J., Liu, D., & Meaney, M. J. (1999). Nongenomic Transmission Across Generations of Maternal Behavior and Stress Responses in the Rat. *Science*, *286*(5442), 1155–1158. https://doi.org/10.1126/science.286.5442.1155

Francis, D. D., Diorio, J., Plotsky, P. M., & Meaney, M. J. (2002). Environmental enrichment reverses the effects of maternal separation on stress reactivity. *Journal of Neuroscience*, *22*(18), 7840-7843.

Frankenhuis, W. E., & Panchanathan, K. (2011). Balancing sampling and specialization: an adaptationist model of incremental development. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1724), 3558-3565.

Frankenhuis, W. E., Nettle, D., & Dall, S. R. X. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 20180110. https://doi.org/10.1098/rstb.2018.0110

Freeman, H. D., & Gosling, S. D. (2010). Personality in nonhuman primates: A review and evaluation of past research. *American Journal of Primatology*, *72*(8), 653–671. https://doi.org/10.1002/ajp.20833

Glover, V., O'Donnell, K. J., O'Connor, T. G., & Fisher, J. (2018). Prenatal maternal stress, fetal programming, and mechanisms underlying later psychopathology—A global perspective. *Development and Psychopathology*, *30*(3), 843–854. https://doi.org/10.1017/S095457941800038X

Gluckman, P. D., Hanson, M. A., & Spencer, H. G. (2005). Predictive adaptive responses and human evolution. *Trends in Ecology & Evolution*, *20*(10), 527–533. https://doi.org/10.1016/j.tree.2005.08.001

Gosling, S. (2001). From Mice to Men: What Can We Learn About Personality From Animal Research? *Psychological Bulletin*, *127*(1), 45–86.

Gros-Louis, J., Perry, S., & Manson, J. H. (2003). Violent coalitionary attacks and intraspecific killing in wild white-faced capuchin monkeys (Cebus capucinus). *Primates*, *44*, 341–346.

Hakeem, A., Sandoval, G. R., & Jones, M. (1996). Brain and life span in primates. In *Handbook of the Psychology of Aging* (Vol. 78).

Haller, J. (2014). The Glucocorticoid/Aggression Relationship in Animals and Humans: An Analysis Sensitive to Behavioral Characteristics, Glucocorticoid Secretion Patterns, and Neural Mechanisms. In *Neuroscience of Aggression: Current Topics in Behavioral Neurosciences* (Vol. 17). Springer. https://link.springer.com/chapter/10.1007/7854_2014_284

Haller, J. (2018). The Role of the Lateral Hypothalamus in Violent Intraspecific Aggression—The Glucocorticoid Deficit Hypothesis. *Frontiers in Systems Neuroscience*, *12*.

Haller, J., Halász, J., Mikics, É., & Kruk, M. R. (2004). Chronic Glucocorticoid Deficiency-Induced Abnormal Aggression, Autonomic Hypoarousal, and Social Deficit in Rats. *Journal of Neuroendocrinology*, *16*(6), 550–557. https://doi.org/10.1111/j.1365-2826.2004.01201.x

Hartwig, W., Rosenberger, A. L., Norconk, M. A., & Owl, M. Y. (2011). Relative Brain Size, Gut Size, and Evolution in New World Monkeys. *The Anatomical Record*, *294*(12), 2207–2221. https://doi.org/10.1002/ar.21515

Harvey, P. H., & Clutton-Brock, T. H. (1985). Life History Variation in Primates. *Evolution*, *39*(3), 559–581. https://doi.org/10.1111/j.1558-5646.1985.tb00395.x

Hau, M., Casagrande, S., Ouyang, J. Q., & Baugh, A. T. (2016). Glucocorticoid-Mediated Phenotypes in Vertebrates. In *Advances in the Study of Behavior* (Vol. 48, pp. 41–115). Elsevier. https://doi.org/10.1016/bs.asb.2016.01.002

Hau, M., Deimel, C., & Moiron, M. (2022). Great tits differ in glucocorticoid plasticity in response to spring temperature. *Proceedings of the Royal Society B: Biological Sciences*, *289*.

Heim, C., & Nemeroff, C. B. (1999). The impact of early adverse experiences on brain systems involved in the pathophysiology of anxiety and affective disorders. *Biological Psychiatry*, *46*(11), 1509–1522. https://doi.org/10.1016/S0006-3223(99)00224-3

Hengartner, M. P. (2017). The Evolutionary Life History Model of Externalizing Personality: Bridging Human and Animal Personality Science to Connect Ultimate and Proximate Mechanisms Underlying Aggressive Dominance, Hostility, and Impulsive Sensation Seeking. *Review of General Psychology*, *21*(4), 330–353. https://doi.org/10.1037/gpr0000127

Hertel, A. G., Niemelä, P. T., Dingemanse, N. J., & Mueller, T. (2020). A guide for studying among-individual behavioral variation from movement data in the wild. *Movement Ecology*, *8*, 30. https://doi.org/10.1186/s40462-020-00216-8

Hill, J. W., & Elias, C. F. (2018). Neuroanatomical Framework of the Metabolic Control of Reproduction. *Physiological Reviews*, *98*(4), 2349–2380. https://doi.org/10.1152/physrev.00033.2017

Hirschenhauser, K., & Oliveira, R. F. (2006). Social modulation of androgens in male vertebrates: Meta-analyses of the challenge hypothesis. *Animal Behaviour*, *71*(2), 265–277. https://doi.org/10.1016/j.anbehav.2005.04.014

Hogan, J. D., Jack, K. M., Campos, F. A., Kalbitzer, U., & Fedigan, L. M. (2019). Group versus population level demographics: An analysis of comparability using long term data on

wild white-faced capuchin monkeys (*Cebus capucinus imitator*). *American Journal of Primatology*, *81*(7), e23027. https://doi.org/10.1002/ajp.23027

Houston, A. I., & McNamara, J. M. (1999). *Models of Adaptive Behaviour: An Approach Based on State*. Cambridge University Press.

Humphreys, K. L., & Zeanah, C. H. (2015). Deviations from the Expectable Environment in Early Childhood and Emerging Psychopathology. *Neuropsychopharmacology*, *40*(1), 154–170. https://doi.org/10.1038/npp.2014.165

Isler, K., & Van Schaik, C. P. (2014). How humans evolved large brains: Comparative evidence. *Evolutionary Anthropology: Issues, News, and Reviews*, *23*(2), 65–75. https://doi.org/10.1002/evan.21403

Jack, K., & Fedigan, L. M. (2004). The Demographic and Reproductive Context of Male Replacements in Cebus Capucinus. *Behaviour*, *141*(6), 755–775. https://doi.org/10.1163/1568539042245178

Jasienska, G., & Ellison, P. T. (2004). Energetic factors and seasonal changes in ovarian function in women from rural Poland. *American Journal of Human Biology*, *16*(5), 563–580. https://doi.org/10.1002/ajhb.20063

Kajokaite, K., Whalen, A., Panchanathan, K., & Perry, S. (2019). White-faced capuchin monkeys use both rank and relationship quality to recruit allies. *Animal Behaviour*, *154*, 161–169. https://doi.org/10.1016/j.anbehav.2019.06.008

Koolhaas, J. M., Korte, S. M., de Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W., & Blockhuis, H. J. (1999). Coping styles in animals: Current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, *23*, 925–935.

Langkilde, T., & Shine, R. (2006). How much stress do researchers inflict on their study animals? A case study using a scincid lizard, Eulamprus heatwolei. *Journal of Experimental Biology*, *209*(6), 1035-1043.

Lea, A. J., Altmann, J., Alberts, S. C., & Tung, J. (2015). Developmental Constraints in a Wild Primate. *The American Naturalist*, *185*(6), 809–821. https://doi.org/10.1086/681016

Lea, A. J., Tung, J., Archie, E. A., & Alberts, S. C. (2017). Developmental plasticity. *Evolution, Medicine, and Public Health*, *2017*(1), 162–175. https://doi.org/10.1093/emph/eox019

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P. M., & Meaney, M. J. (1997). Maternal Care, Hippocampal Glucocorticoid Receptors, and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science*, *277*(5332), 1659–1662. https://doi.org/10.1126/science.277.5332.1659

Lopez-Duran, N., Olson, S., Hajal, N., Felt, B., & Vazquez, D. (2009). Hypothalamic Pituitary Adrenal Axis Functioning in Reactive and Proactive Aggression in Children. *Journal of Abnormal Child Psychology*, *37*, 169–182.

Lu, A., Petrullo, L., Carrera, S., Feder, J., Schneider-Crease, I., & Snyder-Mackler, N. (2019). Developmental responses to early-life adversity: Evolutionary and mechanistic perspectives. *Evolutionary Anthropology: Issues, News, and Reviews*, *28*(5), 249–266. https://doi.org/10.1002/evan.21791

Luevano, L., Sutherland, C., Gonzalez, S. J., & Hernández-Pacheco, R. (2022). Rhesus macaques compensate for reproductive delay following ecological adversity early in life. *Ecology and Evolution*, *12*(1), e8456. https://doi.org/10.1002/ece3.8456

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, *10*(6), 434–445. https://doi.org/10.1038/nrn2639

Luttbeg, B., & Sih, A. (2010). Risk, resources and state-dependent adaptive behavioural syndromes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 3977–3990. https://doi.org/10.1098/rstb.2010.0207

MacLeod, K. J., English, S., Ruuskanen, S. K., & Taborsky, B. (2023). Stress in the social context: A behavioural and eco-evolutionary perspective. *Journal of Experimental Biology*, *226*(15), jeb245829. https://doi.org/10.1242/jeb.245829

Maestripieri, D. (2005). Effects of early experience on female behavioural and reproductive development in rhesus macaques. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1569), 1243–1248. https://doi.org/10.1098/rspb.2005.3059

Maestripieri, D., Higley, J. D., Lindell, S. G., Newman, T. K., McCormack, K. M., & Sanchez, M. M. (2006). Early maternal rejection affects the development of monoaminergic systems and adult abusive parenting in rhesus macaques (Macaca mulatta). *Behavioral Neuroscience*, *120*(5), 1017–1024. https://doi.org/10.1037/0735-7044.120.5.1017

Manson, J. H., & Perry, S. (2013). Personality structure, sex differences, and temporal change and stability in wild white-faced capuchins (Cebus capucinus). *Journal of Comparative Psychology*, *127*(3), 299–311. https://doi.org/10.1037/a0031316

McCormack, K., Newman, T. K., Higley, J. D., Maestripieri, D., & Sanchez, M. M. (2009). Serotonin transporter gene variation, infant abuse, and responsiveness to stress in rhesus macaque mothers and infants. *Hormones and Behavior*, *55*(4), 538–547. https://doi.org/10.1016/j.yhbeh.2009.01.009

McCormack, K., Bramlett, S., Morin, E. L., Siebert, E. R., Guzman, D., Howell, B., & Sanchez, M. M. (2025). Long-Term Effects of Adverse Maternal Care on Hypothalamic–Pituitary–Adrenal (HPA) Axis Function of Juvenile and Adolescent Macaques. *Biology*, *14*(2), 204.

McEwen, B. S. (1998). Stress, Adaptation, and Disease: Allostasis and Allostatic Load. *Annals of the New York Academy of Sciences*, *840*(1), 33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x

McEwen, B. S., & Stellar, E. (1993). Stress and the Individual: Mechanisms Leading to Disease. *Archives of Internal Medicine*, *153*(18), 2093–2101. https://doi.org/10.1001/archinte.1993.00410180039004

McEwen, B. S., & Wingfield, J. C. (2003). The concept of allostasis in biology and biomedicine. *Hormones and Behavior*, *43*(1), 2–15. https://doi.org/10.1016/S0018-506X(02)00024-7

McLaughlin, K. A., & Sheridan, M. A. (2016). Beyond Cumulative Risk: A Dimensional Approach to Childhood Adversity. *Current Directions in Psychological Science*, *25*(4), 239–245. https://doi.org/10.1177/0963721416655883

McLaughlin, K. A., Sheridan, M. A., Humphreys, K. L., Belsky, J., & Ellis, B. J. (2021). The Value of Dimensional Models of Early Experience: Thinking Clearly About Concepts and Categories. *Perspectives on Psychological Science*, *16*(6), 1463–1472. https://doi.org/10.1177/1745691621992346

McNamara, J. M., & Houston, A. I. (1996). State-dependent life histories. *Nature*, *380*(6571), 215–221. https://doi.org/10.1038/380215a0

Meaney, M. J. (2001). Maternal Care, Gene Expression, and the Transmission of Individual Differences in Stress Reactivity Across Generations. *Annual Review of Neuroscience*, *24*(1), 1161–1192. https://doi.org/10.1146/annurev.neuro.24.1.1161

Meyer, J. S., & Hamel, A. F. (2014). Models of Stress in Nonhuman Primates and Their Relevance for Human Psychopathology and Endocrine Dysfunction. *ILAR Journal*, *55*(2), 347–360. https://doi.org/10.1093/ilar/ilu023

Monaghan, P. (2008). Early growth conditions, phenotypic development and environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1497), 1635–1645. https://doi.org/10.1098/rstb.2007.0011

Monaghan, P., & Haussmann, M. F. (2015). The positive and negative consequences of stressors during early life. *Early Human Development*, *91*(11), 643–647. https://doi.org/10.1016/j.earlhumdev.2015.08.008

Murphy, F., Nasa, A., Cullinane, D., Raajakesary, K., Gazzaz, A., Sooknarine, V., Haines, M., Roman, E., Kelly, L., O'Neill, A., Cannon, M., & Roddy, D. W. (2022). Childhood Trauma, the HPA Axis and Psychiatric Illnesses: A Targeted Literature Synthesis. *Frontiers in Psychiatry*, *13*. https://doi.org/10.3389/fpsyt.2022.748372

Nelson, R. J., & Trainor, B. C. (2007). Neural mechanisms of aggression. *Nature Reviews Neuroscience*, *8*(7), 536–546. https://doi.org/10.1038/nrn2174

Nettle, D., & Bateson, M. (2015). Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151005. https://doi.org/10.1098/rspb.2015.1005

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131343. https://doi.org/10.1098/rspb.2013.1343

Onyango, P. O., Gesquiere, L. R., Wango, E. O., Alberts, S. C., & Altmann, J. (2008). Persistence of maternal effects in baboons: Mother's dominance rank at son's conception predicts stress hormone levels in subadult males. *Hormones and Behavior*, *54*(2), 319–324. https://doi.org/10.1016/j.yhbeh.2008.03.002

Palme, R. (2019). Non-invasive measurement of glucocorticoids: Advances and problems. *Physiology & Behavior*, *199*, 229–243. https://doi.org/10.1016/j.physbeh.2018.11.021

Parker, K. J., & Maestripieri, D. (2011). Identifying key features of early stressful experiences that produce stress vulnerability and resilience in primates. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1466–1483. https://doi.org/10.1016/j.neubiorev.2010.09.003

Patterson, S., Hinde, K., Bond, A., Trumble, B., Strum, S., & Silk, J. (2021). Effects of early life adversity on maternal effort and glucocorticoids in wild olive baboons. *Behavioral Ecology and Sociobiology*, *75*(114).

Peckins, M. K., Susman, E. J., Negriff, S., Noll, J., & Trickett, P. K. (2015). Cortisol profiles: A test for adaptive calibration of the stress response system in maltreated and nonmaltreated youth. *Development and Psychopathology*, *27*(4pt2), 1461–1470. https://doi.org/10.1017/S0954579415000875

Pemberton, J. M., Kruuk, L. E., & Clutton-Brock, T. (2022). The unusual value of long-term studies of individuals: The example of the Isle of Rum red deer project. *Annual Review of Ecology, Evolution, and Systematics*, *53*(1), 327-351.

Perry, S. (1997). Male-Female Social Relationships in Wild White-Faced Capuchins (Cebus capucinus). *Behaviour*, *134*(7/8), 477–510.

Perry, S. (1998). Male-Male Social Relationships in Wild White-Faced Capuchins, Cebus capucinus. *Behaviour*, *135*(2), 139–172.

Perry, S. (2011). Social traditions and social learning in capuchin monkeys (Cebus). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 988–996. https://doi.org/10.1098/rstb.2010.0317

Perry, S. (2012). The Behavior of Wild White-Faced Capuchins: Demography, Life History, Social Relationships, and Communication. In *Advances in the Study of Behavior* (Vol. 44, pp. 135–181). Academic Press. https://doi.org/10.1016/B978-0-12-394288-3.00004-6

Perry, S. (2020). Behavioural variation and learning across the lifespan in wild white-faced capuchin monkeys. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1803), 20190494. https://doi.org/10.1098/rstb.2019.0494

Perry, S., Baker, M., Fedigan, L., Gros-Louis, J., Jack, K., MacKinnon, K. C., Manson, J. H., Panger, M., Pyle, K., & Rose, L. (2003). Social Conventions in Wild White-faced Capuchin Monkeys: Evidence for Traditions in a Neotropical Primate. *Current Anthropology*, *44*(2), 241–268. https://doi.org/10.1086/345825

Perry, S., Godoy, I., Lammers, W., & Lin, A. (2017). Impact of personality traits and early life experience on timing of emigration and rise to alpha male status for wild male white-faced capuchin monkeys (Cebus capucinus) at Lomas Barbudal Biological Reserve, Costa Rica. *Behaviour*, *154*, 195–226.

Pritchard, A. J., Capitanio, J. P., Del Rosso, L., McCowan, B., & Vandeleest, J. J. (2023). Hair and plasma cortisol throughout the first 3 years of development in infant rhesus

macaques, Macaca mulatta. *Developmental Psychobiology*, *65*(8), e22437. https://doi.org/10.1002/dev.22437

Pritchard, A. J., & Palombit, R. A. (2022). Measuring behavioral coping style and stress reactivity experimentally in wild olive baboons. *Behavioural Processes*, *195*, 104564. https://doi.org/10.1016/j.beproc.2021.104564

Pryce, C., Aubert, Y., & Maier, C. (2010). The developmental impact of prenatal stress, prenatal dexamethasone and postnatal social stress on physiology, behaviour and neuroanatomy of primate offspring: Studies in rhesus macaque and common marmoset. *Psychopharmacology*, *214*, 33–53.

Qu, J., Fletcher, Q. E., Réale, D., Li, W., & Zhang, Y. (2018). Independence between coping style and stress reactivity in plateau pika. *Physiology & Behavior*, *197*, 1–8. https://doi.org/10.1016/j.physbeh.2018.09.007

Rieger, N. S., Guoynes, C. D., Monari, P. K., Hammond, E. R., Malone, C. L., & Marler, C. A. (2022). Neuroendocrine mechanisms of aggression in rodents. *Motivation Science*, *8*(2), 81–105. https://doi.org/10.1037/mot0000260

Roff, D. (2001). *Life History Evolution* (1st ed.). Sinauer Associates.

Rojas, C. A., Holekamp, K. E., Jasso, M. V., Souza, V., Eisen, J. A., & Theis, K. R. (2022). Taxonomic, genomic, and functional variation in the gut microbiomes of wild spotted hyenas across two decades of study. *bioRxiv*, 2022-08.

Romero, L. M. (2004). Physiological stress in ecology: lessons from biomedical research. *Trends in ecology & evolution*, *19*(5), 249-255.

Rosenbaum, S., Zeng, S., Campos, F. A., Gesquiere, L. R., Altmann, J., Alberts, S. C., Li, F., & Archie, E. A. (2020). Social bonds do not mediate the relationship between early

adversity and adult glucocorticoids in wild baboons. *Proceedings of the National Academy of Sciences*, *117*(33), 20052–20062. https://doi.org/10.1073/pnas.2004524117

Rosvall, K. A. (2013). Life History Trade-Offs and Behavioral Sensitivity to Testosterone: An Experimental Test When Female Aggression and Maternal Care Co-Occur. *PLOS ONE*, *8*(1), e54120. https://doi.org/10.1371/journal.pone.0054120

Sanchez, M. M. (2006). The impact of early adverse care on HPA axis development: Nonhuman primate models. *Hormones and Behavior, Translational Topics in Behavioral Neuroendocrinology*, *50*(4), 623–631. https://doi.org/10.1016/j.yhbeh.2006.06.012

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions*. *Endocrine Reviews*, *21*(1), 55–89. https://doi.org/10.1210/edrv.21.1.0389

Schjolden, J., Backström, T., Pulman, K. G. T., Pottinger, T. G., & Winberg, S. (2005). Divergence in behavioural responses to stress in two strains of rainbow trout (*Oncorhynchus mykiss*) with contrasting stress responsiveness. *Hormones and Behavior*, *48*(5), 537–544. https://doi.org/10.1016/j.yhbeh.2005.04.008

Schoenemann, K. L., & Bonier, F. (2018). Repeatability of glucocorticoid hormones in vertebrates: A meta-analysis. *PeerJ*, *6*, e4398. https://doi.org/10.7717/peerj.4398

Schoenle, L. A., Zimmer, C., Miller, E. T., & Vitousek, M. N. (2021). Does variation in glucocorticoid concentrations predict fitness? A phylogenetic meta-analysis. *General and Comparative Endocrinology*, *300*, 113611. https://doi.org/10.1016/j.ygcen.2020.113611

Seltmann, M. W., Jaatinen, K., Steele, B. B., & Öst, M. (2014). Boldness and Stress Responsiveness as Drivers of Nest-Site Selection in a Ground-Nesting Bird. *Ethology*, *120*(1), 77–89. https://doi.org/10.1111/eth.12181

Shalev, I., & Belsky, J. (2016). Early-life stress and reproductive cost: A two-hit developmental model of accelerated aging? *Medical Hypotheses*, *90*, 41–47. https://doi.org/10.1016/j.mehy.2016.03.002

Sheridan, M. A., & McLaughlin, K. A. (2014). Dimensions of Early Experience and Neural Development: Deprivation and Threat. *Trends in Cognitive Sciences*, *18*(11), 580–585. https://doi.org/10.1016/j.tics.2014.09.001

Sheriff, M. J., Dantzer, B., Delehanty, B., Palme, R., & Boonstra, R. (2011). Measuring stress in wildlife: Techniques for quantifying glucocorticoids. *Oecologia*, *166*(4), 869–887. https://doi.org/10.1007/s00442-011-1943-y

Sheriff, M. J., Krebs, C. J., & Boonstra, R. (2010). The ghosts of predators past: Population cycles and the role of maternal programming under fluctuating predation risk. *Ecology*, *91*(10), 2983–2994. https://doi.org/10.1890/09-1108.1

Sih, A. (2011). Effects of early stress on behavioral syndromes: An integrated adaptive perspective. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1452–1465. https://doi.org/10.1016/j.neubiorev.2011.03.015

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: An ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*(7), 372–378. https://doi.org/10.1016/j.tree.2004.04.009

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

Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: A behavioural ecology perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2762–2772.

Sih, A., Mathot, K. J., Moirón, M., Montiglio, P.-O., Wolf, M., & Dingemanse, N. J. (2015). Animal personality and state–behaviour feedbacks: A review and guide for empiricists. *Trends in Ecology & Evolution*, *30*(1), 50–60. https://doi.org/10.1016/j.tree.2014.11.004

Sloboda, Z., Stephens, R. C., Stephens, P. C., Grey, S. F., Teasdale, B., Hawthorne, R. D., Williams, J., & Marquette, J. F. (2009). The Adolescent Substance Abuse Prevention Study: A randomized field trial of a universal substance abuse prevention program. *Drug and Alcohol Dependence*, *102*(1), 1–10. https://doi.org/10.1016/j.drugalcdep.2009.01.015

Speranza, L., Filiz, K. D., Lippiello, P., Ferraro, M. G., Pascarella, S., Miniaci, M. C., & Volpicelli, F. (2024). Enduring Neurobiological Consequences of Early-Life Stress: Insights from Rodent Behavioral Paradigms. *Biomedicines*, *12*(9), 1978. https://doi.org/10.3390/biomedicines12091978

Stamps, J. (2003). Behavioural processes affecting development: Tinbergen's fourth question comes of age. Animal Behaviour 66, 1–13

Stamps, J. A. (2007). Growth-mortality tradeoffs and 'personality traits' in animals. *Ecology Letters*, *10*(5), 355–363. https://doi.org/10.1111/j.1461-0248.2007.01034.x

Stamps, J., & Groothuis, T. G. G. (2010). The development of animal personality: Relevance, concepts and perspectives. *Biological Reviews*, *85*(2), 301–325. https://doi.org/10.1111/j.1469-185X.2009.00103.x

Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259–268. https://doi.org/10.2307/2389364

Stearns, S. C. (2000). Life history evolution: Successes, limitations, and prospects. *Naturwissenschaften*, *87*(11), 476–486. https://doi.org/10.1007/s001140050763

Stibor, H. (1992). Predator induced life-history shifts in a freshwater cladoceran. *Oecologia*, *92*(2), 162–165.

Strauss, E. D., Shizuka, D., & Holekamp, K. E. (2020). Juvenile rank acquisition is associated with fitness independent of adult rank. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1922), 20192969. https://doi.org/10.1098/rspb.2019.2969

Szyf, M., McGowan, P., & Meaney, M. J. (2008). The social environment and the epigenome. *Environmental and Molecular Mutagenesis*, *49*(1), 46–60. https://doi.org/10.1002/em.20357

Taborsky, B. (2016). Opening the black box of developmental experiments: Behavioural mechanisms underlying long-term effects of early social experience. *Ethology*, *122*(4), 267-283.

Taff, C. C., Schoenle, L. A., & Vitousek, M. N. (2018). The repeatability of glucocorticoids: A review and meta-analysis. *General and Comparative Endocrinology*, *260*, 136–145. https://doi.org/10.1016/j.ygcen.2018.01.011

Thomson, J. S., Watts, P. C., Pottinger, T. G., & Sneddon, L. U. (2011). Physiological and genetic correlates of boldness: Characterising the mechanisms of behavioural variation in rainbow trout, *Oncorhynchus mykiss*. *Hormones and Behavior*, *59*(1), 67–74. https://doi.org/10.1016/j.yhbeh.2010.10.010

Tollrian, R. (1995). Predator-Induced Morphological Defenses: Costs, Life History Shifts, and Maternal Effects in Daphnia Pulex. *Ecology*, *76*(6), 1691–1705. https://doi.org/10.2307/1940703

Touma, C., & Palme, R. (2005). Measuring Fecal Glucocorticoid Metabolites in Mammals and Birds: The Importance of Validation. *Annals of the New York Academy of Sciences*, *1046*(1), 54–74. https://doi.org/10.1196/annals.1343.006

Tung, J., Archie, E. A., Altmann, J., & Alberts, S. C. (2016). Cumulative early life adversity predicts longevity in wild baboons. *Nature Communications*, *7*(1), 11181. https://doi.org/10.1038/ncomms11181

Uher, J. (2011a). Individual behavioral phenotypes: An integrative meta-theoretical framework. Why "behavioral syndromes" are not analogs of "personality." *Developmental Psychobiology*, *53*(6), 521–548. https://doi.org/10.1002/dev.20544

Uher, J. (2011b). Personality in Nonhuman Primates: What Can We Learn from Human Personality Psychology? In A. Weiss, J. E. King, & L. Murray (Eds.), *Personality and Temperament in Nonhuman Primates* (pp. 41–76). Springer. https://doi.org/10.1007/978-1-4614-0176-6_3

van Goozen, S. H. M., Fairchild, G., Snoek, H., & Harold, G. T. (2007). The evidence for a neurobiological model of childhood antisocial behavior. *Psychological Bulletin*, *133*(1), 149–182. https://doi.org/10.1037/0033-2909.133.1.149

Van Honk, J., Harmon-Jones, E., Morgan, B. E., & Schutter, D. J. L. G. (2010). Socially Explosive Minds: The Triple Imbalance Hypothesis of Reactive Aggression. *Journal of Personality*, *78*(1), 67–94. https://doi.org/10.1111/j.1467-6494.2009.00609.x

van Schaik, C. P., & Isler, K. (2012). Life-History Evolution in Primates. In *The Evolution of Primate Societies*. University of Chicago Press.

Vitzthum, V. J. (2008). Evolutionary Models of Women's Reproductive Functioning. *Annual Review of Anthropology*, *37*(1), 53–73. https://doi.org/10.1146/annurev.anthro.37.081407.085112

Waters, R. C., & Gould, E. (2022). Early Life Adversity and Neuropsychiatric Disease: Differential Outcomes and Translational Relevance of Rodent Models. *Frontiers in Systems Neuroscience*, *16*, 860847. https://doi.org/10.3389/fnsys.2022.860847

Watts, D. P. (2011). Long-term research on chimpanzee behavioral ecology in Kibale National Park, Uganda. In *Long-term field studies of primates* (pp. 313-338). Berlin, Heidelberg: Springer Berlin Heidelberg.

Weaver, I. C. g., Diorio, J., Seckl, J. R., Szyf, M., & Meaney, M. J. (2004). Early Environmental Regulation of Hippocampal Glucocorticoid Receptor Gene Expression: Characterization of Intracellular Mediators and Potential Genomic Target Sites. *Annals of the New York Academy of Sciences*, *1024*(1), 182–212. https://doi.org/10.1196/annals.1321.099

Weaver, I. C. G., Meaney, M. J., & Szyf, M. (2006). Maternal care effects on the hippocampal transcriptome and anxiety-mediated behaviors in the offspring that are reversible in adulthood. *Proceedings of the National Academy of Sciences*, *103*(9), 3480–3485. https://doi.org/10.1073/pnas.0507526103

Weibel, C. J., Tung, J., Alberts, S. C., & Archie, E. A. (2020). Accelerated reproduction is not an adaptive response to early-life adversity in wild baboons. *Proceedings of the National Academy of Sciences*, *117*(40), 24909–24919. https://doi.org/10.1073/pnas.2004018117

Weinstock, M. (2008). The long-term behavioural consequences of prenatal stress. *Neuroscience & Biobehavioral Reviews, The Long-Term Consequences of Stress on Brain Function:*

From Adaptation to Mental Diseases, *32*(6), 1073–1086. https://doi.org/10.1016/j.neubiorev.2008.03.002

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.

Westrick, S., van Kesteren, F., Palme, R., Boonstra, R., Lane, J., Boutin, S., McAdam, A., & Dantzer, B. (2019). Stress activity is not predictive of coping style in North American red squirrels. *Behavioral Ecology and Sociobiology*, *73*(113).

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

Wolf, M., & Weissing, F. J. (2012). Animal personalities: Consequences for ecology and evolution. *Trends in Ecology & Evolution*, *27*(8), 452–461. https://doi.org/10.1016/j.tree.2012.05.001

Wood, E. K., Aston, S. A., O'Connell, P. H., Hafen, E., Skowbo, A. N., Schwandt, M. L., ... & Higley, J. D. (2024). Genotypic variation in the promoter region of the CRH-248 gene interacts with early rearing experiences to disrupt the development of the HPA axis in infant rhesus macaques (Macaca mulatta). *Stress*, *27*(1), 2377272.

Zipple, M. N., Archie, E. A., Tung, J., Altmann, J., & Alberts, S. C. (2019). Intergenerational effects of early adversity on survival in wild baboons. *eLife*, *8*, e47433. https://doi.org/10.7554/eLife.47433

Chapter 1: Early-Life Adversity Shapes Mortality and Reproductive Strategies in Wild Female Capuchin Monkeys

Abstract

Adverse environmental conditions during early development can be a critical determinant of health, behavior, and life history outcomes across the lifespan. It is widely accepted that early life adversity (ELA) leads to a host of altered biological and behavioral processes, many of which can place organisms at increased risk for later-stage dysfunction and disease. Evolutionary models also propose that ELA acts as a cue of future conditions, leading to adaptive plasticity in reproductive scheduling. Predictive models hypothesize that ELA signals high extrinsic mortality, promoting an adaptive "live fast, die young" strategy characterized by accelerated reproduction. Conversely, constraint models posit that ELA imposes physiological limitations, necessitating a delayed reproductive onset when metabolic energy is insufficient to meet the high energetic and parental costs of reproduction. We tested these competing hypotheses by examining the impact of ELA on mortality and reproductive life history trajectories in wild, female white-faced capuchin monkeys. We analyzed the effects of distinct stressors across three developmental windows: late prenatal, first year, and first five years of life. From-birth survival models revealed that adversity timing and type strongly predicted mortality risk with the largest effects concentrated in prenatal psychosocial adversity, alpha male replacement during the prenatal period and maternal loss during the first year. In contrast, delayed-entry models conditioning on survival to age five show no residual mortality effects, indicating that ELA's lethality is concentrated early in life. Among females who survived to the core reproductive period, psychosocial adversity predicted earlier age at first birth, but this effect was age dependent. Ecological/energetic adversity during development showed more limited associations

with later reproductive performance, including offspring survival, birth spacing, and age-expected fertility. Together, these findings suggest that reproductive acceleration, developmental constraint, and viability selection may contribute differently to life-history variation depending on the type and timing of adversity, underscoring the importance of distinguishing between adversity types.

Introduction

Life history theory provides a fundamental framework for understanding how organisms allocate finite metabolic energy among the competing demands of maintenance, growth, and reproduction (Stearns, 1976). Because energetic investment in one domain often limits investment in another, organisms must navigate trade-offs that shape patterns of development, reproduction, and survival (Stearns, 1976, 1989). Ecological conditions influence the evolution of these trade-offs, contributing to substantial variation in reproductive traits across and within species (Charnov & Berrigan, 1993; Harvey & Clutton-Brock, 1985; Roff, 2001; Stearns, 1989). Developmental plasticity can further modify trait expression in response to local conditions, particularly during sensitive periods (West-Eberhard, 2003). In long-lived social species, early ecological and social conditions may be especially consequential, shaping stress physiology, behavioral development, and later-life condition through pathways that link early experience to somatic state and life-history trajectories (Lea et al., 2015; Tung et al., 2016).

Conditions experienced early in life are therefore expected to influence later physiological and reproductive trajectories. Among the many environmental influences encountered during development, some experiences may impose substantial energetic, physiological, or social challenges during sensitive developmental windows (Brown et al., 2024; Ellis et al., 2009; Lea et al., 2017; Lu et al., 2019; Nettle et al., 2013; Urlacher, 2023). Within the

evolutionary and comparative literature, these experiences are often discussed under the broader concept of early-life adversity (ELA). Comparative and evolutionary-developmental frameworks propose that early environmental conditions can shape behavioral, physiological, and reproductive development through developmental plasticity, although whether such responses function as predictive adaptive responses remains difficult to test, particularly in long-lived primates (Brown et al., 2024; Ellis et al., 2009; Lea et al., 2017; Lu et al., 2019; Nettle et al., 2013; Urlacher, 2023). At the same time, developmental experiences associated with adversity may also alter somatic condition and increase cumulative physiological costs over time, potentially constraining later-life trajectories (Monaghan, 2008; Taylor et al., 2011). Following the comparative primate literature, we use the term ELA to refer to environmental conditions or developmental experiences that threaten survival, growth, or normal physiological function (Rosenbaum et al., 2025; Tung et al., 2016; Zipple et al., 2019). Across both human and non-human animals, ELA has been associated with downstream effects on health, behavior, and life-history outcomes. In humans, these experiences have been linked to altered behavioral development, increased morbidity, and shortened lifespan (Taylor et al., 2011). Within evolutionary-developmental perspectives, however, ELA has also been interpreted as a source of information about future environmental conditions that may shape developmental trajectories in potentially adaptive ways (Nettle et al., 2013).

However, the interpretation of ELA within predictive adaptive frameworks remains controversial, particularly in long-lived, slow-reproducing species such as humans and other primates (Lea et al., 2017; Lu et al., 2019; Wells, 2007). A central theoretical question is whether environmental conditions encountered early in life can reliably predict ecological or social conditions decades later, or whether substantial environmental unpredictability across long

lifespans limits the utility of predictive developmental responses (Frankenhuis et al., 2019; Nettle & Bateson, 2015). If developmental conditions are unlikely to forecast future external environments, early adversity may be better understood as a source of physiological constraint or as information about the individual's developing somatic state, with important consequences for later-life trajectories (Bateson et al., 2014; Lea et al., 2017; Monaghan, 2008; Nettle & Bateson, 2015). These competing, and sometimes overlapping, interpretations have motivated several evolutionary-developmental frameworks, especially predictive adaptive response, and developmental constraint models. Predictive adaptive response models propose that organisms use early-life conditions as informational cues about future environments, potentially favoring accelerated reproductive development under conditions associated with elevated mortality risk or environmental harshness (Gluckman & Hanson, 2004; Lea et al., 2017; Nettle & Bateson, 2015). Constraint-based models, in contrast, emphasize the direct physiological and energetic costs of adversity, arguing that developmental stressors may impair somatic condition and thereby delay, suppress, or otherwise constrain reproductive development (Nettle & Bateson, 2015; Weibel et al., 2020; Zipple et al., 2019).

In predictive adaptive response models, early-life adversity is hypothesized to function as a cue of elevated extrinsic mortality risk, favoring earlier reproduction in environments where long-term survival is uncertain. Under this framework, organisms may accelerate maturation or reproduction to increase the probability of reproducing before death (Lea et al., 2017; Nettle & Bateson, 2015). Patterns consistent with accelerated reproductive scheduling have been reported in several relatively fast-lived systems, where shortened developmental schedules may plausibly increase fitness under conditions of elevated mortality risk. Experimental studies in rodents, for example, have shown that early-life stress and social instability can accelerate pubertal timing

and reproductive maturation (Bateson et al., 2014). In rhesus macaques, cross-fostering work has similarly found that females reared by mothers providing harsher or more inconsistent care reproduced earlier than females reared under more stable maternal conditions (Maestripieri, 2005). Together, these findings suggest that adverse developmental conditions can, in some contexts, shift reproductive scheduling toward earlier maturation and reproduction, although such patterns do not by themselves establish that these shifts are adaptive responses to predicted future environments.

Evidence from long-lived primates, however, has often been less consistent with predictions of reproductive acceleration. In species where fitness depends heavily on prolonged survival and sustained somatic maintenance, adverse developmental conditions frequently appear to constrain, rather than accelerate, reproductive trajectories (Lea et al., 2015; Petersen et al., 2026). For example, in wild female baboons, cumulative early-life adversity predicted dramatically reduced adult survival: females exposed to three or more sources of adversity had a predicted median survival of 8.85 years, compared with 18.60–24.02 years for females exposed to zero or one source of adversity (Tung et al., 2016). Yet direct tests of predictive adaptive response models in this population found little evidence that adversity accelerated reproductive scheduling or enhanced fertility under harsh adult conditions (Rosenbaum et al., 2025). Instead, early adversity appeared to reduce somatic condition and constrain growth, survival, or reproduction regardless of adult environmental conditions (Lea et al., 2015; Rosenbaum et al., 2025; Tung et al., 2016). Similar patterns have been reported in other long-lived mammals. In Asian elephants, for instance, stressful developmental conditions were associated with accelerated reproductive senescence and a shortened reproductive lifespan (Mumby et al., 2015). Collectively, these findings suggest that in long-lived species, the energetic and physiological

costs of adversity may often constrain later-life reproduction rather than produce the reproductive acceleration predicted by some adaptive developmental models.

Consistent with these findings, constraint models propose that early-life adversity operates primarily through physiological and energetic limitation rather than adaptive prediction, resulting in delayed or suppressed reproductive development (Monaghan, 2008; Uller et al., 2020; Weibel et al., 2020). Energetic models of reproduction emphasize that reproductive function is regulated through trade-offs among growth, maintenance, somatic repair, and reproductive investment (Ellison, 2003; Pontzer, 2015). When energetic resources are limited, whether due to nutritional scarcity, illness, developmental stress, or elevated maintenance demands, organisms may delay or downregulate reproductive processes in order to prioritize survival and somatic condition (Ellison, 2003; Emery Thompson, 2017; Stearns, 1989). Evidence from mammalian reproductive ecology, including humans and other primates, indicates that energetic limitation can alter reproductive maturation, suppress fertility, and constrain reproductive output, particularly in species characterized by slow development and prolonged investment in somatic maintenance (Emery Thompson, 2017; Jasienska & Ellison, 2004; Pontzer, 2015). In long-lived primates, adverse developmental conditions may therefore constrain growth and gonadal maturation by limiting the energy available for somatic development, ultimately delaying reproductive onset and reducing later-life reproductive capacity (Emery Thompson, 2017; Lea et al., 2015; Lu et al., 2019).

The difficulty of distinguishing between predictive and constraint-based processes highlights the need for greater precision in how early-life adversity is conceptualized and measured. Although most dimensional models of adversity have been developed in human developmental research, they highlight a broader measurement problem that is directly relevant

to comparative studies of ELA. A common approach in the ELA literature has been to aggregate diverse experiences into cumulative risk indices, an approach that is useful for identifying broad patterns of exposure but may obscure distinct causal pathways by implicitly treating different forms of adversity as equivalent and additive (McLaughlin et al., 2021; McLaughlin & Sheridan, 2016). In response to this limitation, dimensional models of adversity have sought to characterize early experiences according to their underlying features rather than treating adversity as a unitary construct (Ellis et al., 2009; Humphreys & Zeanah, 2015; McLaughlin et al., 2021; McLaughlin & Sheridan, 2016). These frameworks distinguish experiences involving threat or harm from those involving deprivation or the absence of expected environmental inputs. Increasing evidence from human developmental research suggests that these dimensions are associated with partially distinct developmental, neuroendocrine, and neurobiological pathways (McLaughlin et al., 2021; McLaughlin & Sheridan, 2016; Sheridan & McLaughlin, 2014), raising the possibility that different forms of adversity may also shape life-history trajectories through different mechanisms.

Related evolutionary-developmental frameworks have similarly emphasized that environmental adversity is multidimensional. One influential distinction separates harshness, defined as extrinsic sources of morbidity and mortality, from unpredictability, defined as stochastic temporal variability in morbidity and mortality (Ellis et al., 2009). Under predictive adaptive response models, environments characterized by high harshness or unpredictability are hypothesized to favor accelerated life-history strategies because they reduce the expected benefits of long-term somatic investment (Ellis et al., 2009; Nettle & Bateson, 2015). In contrast, stable or less hazardous environments are expected to favor delayed reproduction and greater investment in maintenance and future reproduction. Importantly, however, sources of adversity

in natural populations rarely map cleanly onto single theoretical dimensions. The same developmental exposure may simultaneously involve energetic limitation, social instability, mortality risk, and environmental unpredictability, complicating efforts to distinguish between competing developmental pathways and life-history responses.

An additional challenge in understanding early-life adversity is that the biological consequences of environmental exposures may depend not only on their type, but also on when they occur during development. Developmental plasticity is often expressed through sensitive windows - periods during which environmental inputs exert disproportionately strong and lasting effects on physiology, behavior, and life-history trajectories (English & Barreaux, 2020; Fawcett & Frankenhuis, 2015). Because developmental systems mature on different schedules, adversity experienced at different stages may influence organisms through partially distinct mechanisms. In long-lived primates, where development is prolonged and reproductive maturation occurs slowly (Kappeler & Pereira, 2003; Leigh, 2004), understanding the timing of adversity may therefore be as important as understanding its source.

Furthermore, sources of adversity in natural populations rarely map cleanly onto single theoretical dimensions or apply exclusively to a particular developmental mechanism; the same developmental exposure may simultaneously involve energetic limitation, social instability, mortality risk, and environmental unpredictability, complicating efforts to distinguish among competing developmental pathways and life-history responses. Although we make broad strokes distinctions between energetic/ecological adversity variables (rainfall, group mortality, group size) and psychosocial variables (alpha male replacement, fission, maternal demise), these are, of course, not completely mutually exclusive in their effects. For example, rainfall primarily impacts nutritional intake, and alpha male replacements primarily increase social stress, by

eliminating some important relationships and restructuring existing coalitionary structures. But in primates, there are often complex feedback relationships between resource abundance and distribution and aspects of social relationships (such as group size and coalition formation). So whereas some of the variables we term primarily psychosocial are primarily of a psychosocial nature, affecting survivorship and reproduction via their impact on important social relationships (Kajokaite et al., 2019; Perry, 2012), they do not have non-zero implications for nutritional intake.

The Present Study

In this study, we use longitudinal life history data from a wild population of white-faced capuchins (*Cebus imitator*) to examine how early-life adversity shapes mortality and reproductive trajectories across the lifespan. We focus on multiple forms of adversity relevant to wild primates, including social instability, maternal loss, demographic instability, ecological variation, and group-level social conditions. Rather than aggregating these exposures into a single cumulative index, we examine how specific forms of adversity operate across different stages of development. Because developmental timing may strongly influence how organisms respond to environmental stressors, we analyze these exposures across three biologically distinct developmental windows: the late prenatal period, the first year of life, and the first five years of life.

We leverage the concept of sensitive windows in development - periods where environmental experience exerts a uniquely strong influence on the resulting phenotype (English & Barreaux, 2020). From an evolutionary perspective, natural selection favors sensitive windows because they represent the optimal temporal trade-off between the reliability of incoming environmental cues and the long-term advantages of phenotypic specialization (Fawcett &

Frankenhuis, 2015). In long-lived species like primates, understanding how environmental conditions interact with the organism's life history requires examining these specific developmental milestones.

The prenatal period represents a well-established sensitive window for developmental plasticity in mammals, during which maternal stress and energetic condition can exert lasting effects on offspring physiology and development (Welberg & Seckl, 2001). Experimental studies in both rodents and laboratory primates have shown that prenatal stress exposure alters offspring HPA axis function and elevates glucocorticoid activity later in life (Clarke et al., 1994; Pryce et al., 2010). Prenatal adversity may also influence offspring phenotype through maternal energetic condition and placental nutrient transfer, contributing to fetal growth restriction, altered metabolic regulation, and reduced somatic condition later in life (Blackmore & Ozanne, 2015; Cottrell, 2009; Herring et al., 2018). In white-faced capuchins, which have a gestational period of approximately 157 ± 8 days (Carnegie et al., 2011), this developmental window corresponds to approximately the final 50 days prior to birth.

The first year of life encompasses the period of immediate postnatal dependency and rapid somatic growth. In altricial primates such as capuchins, the mother serves as both the primary nutritional source and the central social relationship during early development (Godoy et al., 2024). This period is also characterized by high energetic demand, as infant growth and lactational transfer represent major components of reproductive energy allocation in primates (Key & Ross, 1999). Although capuchin infants may continue nursing into the second year of life, maternal investment is especially intensive during the first year, when offspring remain heavily dependent on lactation and maternal care. Adverse experiences during this developmental window, including maternal illness or impending maternal loss, are therefore

likely to disrupt both energetic provisioning and social stability, and are strongly associated with elevated mortality risk (Zipple et al., 2021).

Finally, the window spanning the first five years of life captures the cumulative effect of stressors sustained over the entire period leading up to the onset of reproductive maturity. This developmental window is particularly important in long-lived primates because it encompasses sustained somatic growth, social development, and the gradual accumulation of energetic and physiological costs. In baboons and other long-lived mammals, cumulative early-life adversity has been associated with substantial reductions in adult survival and long-term somatic quality (Tung et al., 2016). Because this period extends across much of juvenile development, adversity experienced during it may be especially likely to influence traits related to future reproductive capacity, including the timing of reproductive onset and the ability to sustain reproduction across adulthood. Rather than accelerating reproduction, cumulative adversity during this window may therefore operate primarily through developmental constraint, limiting somatic investment and reducing later-life reproductive potential.

In this study we aim to test the following hypotheses:

Hypothesis 1: We predict that the effects of early-life adversity on mortality risk will vary across developmental windows, reflecting differences in biological vulnerability across stages of development. Because prenatal development, infancy, and the prolonged juvenile period involve distinct energetic, physiological, and social demands, exposures occurring during these windows are expected to show different associations with later-life survival.

Hypothesis 2: In contrast to predictive adaptive response models, which propose that early-life adversity accelerates reproductive development in response to elevated mortality risk, we predict that adversity in this population will operate primarily through developmental

constraint. Specifically, we predict that greater exposure to adversity during development will be associated with delayed age at first birth, reduced reproductive lifespan, longer interbirth intervals, lower age-expected fertility, and reduced offspring survival.

Methods

All data were collected in accordance with protocols approved by the UCLA ARC/IACUC (protocols 1996-122, 2005-084 & 2016-022). This study adheres to the American Society of Primatologists (ASP) Principles for the Ethical Treatment of Non-Human Primates and Code of Best Practices for Field Primatology. All research activities complied with Costa Rican law.

Study System and Data Collection

This study was conducted on a wild population of white-faced capuchins (*Cebus imitator*) living within a tropical dry forest at the Lomas Barbudal Biological Reserve and adjacent private lands in Guanacaste Province, Costa Rica (10° 30'24" N, 85° 22'17"W elevation ~170 feet). This ecosystem is characterized by pronounced seasonality, with a distinct 5 -month dry season (typically mid-December through mid-May) and a rainy season. Annual rainfall ranges between 500 and 2,500 mm (Stan et al., 2020), exhibiting high interannual variability, closely tied to large-scale climatic patterns like the El Nino Southern Oscillation (ENSO) (Babcock et al., 2016). This seasonality drives a cyclical pattern of resource availability. Phenological cycles in tropical dry forests are often tightly linked to seasonal rainfall, though the timing of key events can vary across sites. At Lomas Barbudal, canopy flushing and flowering are most prominent early in the rain season, while fruiting tends to peak toward the end of the dry season and into the early wet months -- a pattern consistent with observations from other Neotropical dry forests

where fruiting is bimodal, with a major peak at the dry season's end (Detto et al., 2018). This seasonal ebb in resource availability, particularly during the mid-to-late dry season, represents a recurrent source of ecological and energetic stress for omnivorous foragers like capuchins (Carnegie et al., 2011). **Climatic perturbations**, such as droughts or exceptionally wet periods, can therefore have acute and measurable impacts on nutritional intake, making rainfall deviation a biologically meaningful proxy for ecological adversity.

Capuchins are long-lived, arboreal primates living in multi-male, multi-female social groups (Perry, 2012). Their social and ecological dynamics provide naturally occurring, quantifiable stressors that map onto the core dimensions of adversity identified in the previously discussed theoretical frameworks. Capuchin social groups are characterized by female philopatry and male dispersal, with high reproductive skew among males (Perry, 2012). Alpha males sire the majority of offspring, and **alpha male replacement events** (often violent 'takeovers') are a major source of social instability (Jack & Fedigan, 2004). These takeovers are frequently followed by infanticide (Jack & Fedigan, 2004; Perry, 2012), representing a direct and severe threat to infants and a source of chronic psychosocial stress for pregnant females and group members. Furthermore, group **fission events** (wherein the group permanently splits along matrilines (Perry, 2012) cause significant disruption to social networks and relationships. Both takeovers and fissions create environments of unpredictability and social conflict, and perceived threat, which we capture through direct behavioral measures and through a measure of **group mortality rate** - a composite index of the hazardous social and ecological environment experienced by a developing individual.

As omnivores in a seasonal tropical dry forest, capuchins' energetic status is tightly linked to food availability, which is strongly influenced by rainfall. Periods of abnormally high

or low rainfall (climatic perturbations) directly affect the abundance of key resources like fruits (Detto et al., 2018) and arthropods (Newell et al., 2023; Palacios-Vargas et al., 2007), serving as a source of ecological/energetic stress. Additionally, an individual's social **group size** may also be a source of energetic stress insofar as larger groups face increased within-group feeding competition (Tinsley Johnson, 2021; Jacobson et al., 2025), while smaller groups face higher predation and infanticide risk (Borries et al., 2008; Tinsley Johnson et al., 2021). **Maternal loss**, especially before weaning, represents a catastrophic energetic deficit for dependent offspring. The clear signatures of these stressors, combined with the species' slow life history (mean age at first birth between 6.2 and 6.5 years (Fedigan & Jack, 2011; Perry, 2012), interbirth interval ~2 years (Perry, 2012) make this system well suited for the tracking of how specific, well-defined early-life adversities influence mortality and reproductive trajectories across the lifespan.

The Lomas Barbudal Monkey Project (LBMP) has collected near-continuous behavioral and demographic data using non-invasive means on the capuchins of Lomas Barbudal since 1990 (Perry et al, 2012). The dataset used in this study leverages long-term demographic records, including birth and death dates known with a six-month margin of error, group compositions, and life history events for 779 known individuals. Behavioral data collected via focal animal follows and ad libitum data are used to construct metrics of social dynamics, including alpha male replacement events and takeovers. This study includes data collected from 1990 to 2026.

Measures

Detailed operational definitions for all variables are provided in the Supplementary Information. Briefly, we measured early-life adversity and female life history outcomes as follows:

Early-Life Adversity (ELA) Measures.

Climatic perturbations/Rainfall deviations: Climatic perturbations, that is periods of abnormally high or low rainfall and/or temperature, are likely to impact the availability of nutritional resources within the environment serving as source of energetic/metabolic stress. We use data extracted from ERA5 in place of weather station data due to gaps in coverage associated with the available local weather station data. Rainfall deviation was calculated by converting monthly precipitation totals to z-scores relative to the 1979-1989 baseline mean and SD for each calendar month, then averaging these z-scores across each developmental window (see Supplementary Information for additional details.

Social Group Size: The average number of individuals in the social group was calculated for each period. We used census data to sum the total number of coresidents on each day within the period of interest for each individual and then divided this sum by the number of days in the period. As both larger and smaller groups are associated with increased adversity (Tinsley-Johnson et al., 2021), group size was first centered at the mean and squared to allow for modeling of a non-linear effect.

Group Mortality Rate: We calculated the annualized mortality rate within each individual's social group during each developmental window using census data to tally co-resident days. Adult males (≥ 5 years old) were excluded from both the numerator and denominator because, as the dispersing sex, their disappearance from the group cannot reliably be distinguished from emigration and would thus bias mortality estimates. The mortality rate was computed as the total number of co-resident days (excluding adult males) for each day within the developmental window, then multiplied by 365 to produce an annualized rate.

Alpha Male Replacements/Takeovers: The count of alpha male replacements (including takeovers) within the social group was recorded for each period. In brief, changes in the alpha male position were recorded using behavioral observations and census records.

Maternal Demise: We measure maternal loss with a binary variable indicating whether an individual's mother died within the first year or the first five years of life.

Fission Events: Fission events are recorded on the basis of census records and indicated with a binary variable.

Life History Outcome Variables

Age at First Birth: The minimum verified age at first reproduction for females who reached 4.5 years of age (the youngest birth in this population was observed in a female aged 4.9 years). Individuals were censored if they died before reproducing or were alive without a birth by the most recent observation date for their social group.

Reproductive Lifespan: The difference between a female's age at last birth and age at first birth. This variable defines the window of opportunity for reproduction and is a key driver of variation in lifetime reproductive output (Snyder & Ellner, 2024).

Average Interbirth Interval (IBI): The mean number of days between consecutive births for each female, calculated using intervals with reliable observation and regardless of the survival outcome of the previous infant. While short IBIs can follow early infant loss, such loss may itself reflect constraints on maternal investment, making these intervals informative of overall reproductive scheduling and maternal condition. When considered alongside offspring survival (proportion surviving to age 4), IBI helps evaluate potential trade-offs between offspring quantity and quality (Walker et al., 2008).

Proportion of Offspring Surviving: The proportion of a female's offspring that was confirmed to survive to four years of age. This variable assesses the success of maternal investment and captures the "quality" component of the quantity-quality trade-off.

Age-Specific Expected Fertility: The difference between a female's observed number of births and the number expected based on population-level, age-specific fertility rates. This metric evaluates whether (and to what degree) an individual's reproductive output deviates from the population norm.

Statistical Analyses

Mortality Risk: Delayed-Entry Models

We modeled mortality risk using Cox proportional hazards (PH) models with delayed entry (left-truncation) at age five. By conditioning on survival to age five, these models estimate the residual association between early adversity and subsequent mortality risk, disregarding the impact of any direct early-life lethality. The five-year threshold was selected because it falls after the period of infanticide vulnerability (which is concentrated in the first two years), after nutritional dependence on the mother has ended, and before the median age at first birth, separating the exposure period from the outcome period assessed in a parallel age at first birth model (discussed below). Analyses were restricted to females because male emigration creates competing risks that confound mortality estimation: male disappearance from the study population may reflect either death or successful dispersal to an unstudied group, and these events cannot always be distinguished. The primary female sample comprised 105 individuals who survived to age five, of whom 58 died during the observation period and 47 were right censored at the end of the study or at last observation. We fit three separate models corresponding to the three developmental windows, each including the full set of adversity predictors for that window. The proportional

hazards assumption was tested using Schoenfeld residual tests; likelihood ratio tests (LRT) assessed overall model significance (Tables S1.5–S1.6).

Mortality Risk: From-Birth Models

To assess the extent to which delayed entry obscures early-life lethality, we fit parallel Cox PH models without left-truncation, following all females from birth. These from-birth models estimate the total association between each adversity type and mortality across the full lifespan, including the early period in which adversity effects are hypothesized to be strongest. Proportional hazards violations were examined because the effects spanning the developmental period and later life may vary with age. Where the PH assumption was violated, hazard ratios were interpreted as time-varying averaged effects (Tables S1.2-S1.4)

Age at First Birth (AFB): Primary and Sensitivity Models

AFB was modeled using Cox PH models among females surviving to the biological minimum age for reproduction. Primary models used an entry-age at 4.5-year, treating first birth as the event and censoring females who had not yet given birth at the end of the observation period or who died before reproducing. Sensitivity analyses used a six-year threshold. As the 4.5-year threshold captures the earliest physiologically possible age at reproduction, it is likely to include a substantial number of females who are not yet physiologically prepared for mating, potentially diluting programming signals detectable only among females who have clearly entered the effective reproductive window. The six-year threshold was chosen to approximate the median age at first birth in this population (approximately 6.2 years), restricting the sample to females who have entered the core reproductive period (Tables S1.12–S1.13). We treat the 4.5-year threshold as the primary specification because it is anchored to the youngest possible age associated with a

reproductive event observed in this population and because restricting to survivors of an older threshold would itself discard information on the youngest-reproducing females. The six-year threshold was included as a sensitivity analysis focused on females who had reached the core reproductive period, recognizing that restricting the sample to females whose first birth occurred no earlier than six years necessarily excludes females who reproduced earlier.

Statistical Analyses: Regression Models

Multiple linear regression models were used to examine associations between ELA measures and life history outcomes among parous females. Twelve models were fit in total (four outcomes - reproductive lifespan, IBI, offspring survival to age four, and age-expected fertility - by three developmental windows). False discovery rate (FDR) correction was applied using the Benjamini-Hochberg procedure at two levels: within predictor families and across all predictors within each developmental window. Within-family correction is reported as the primary result, with window-wide correction provided as a more conservative sensitivity check. Model fit statistics including R-squared, adjusted R-squared, and overall F-test results with associated p-values are reported for all twelve models (Tables S1.14–S1.16).

All statistical analyses were conducted in R version 4.3.1 (R Core Team, 2023) using the survival package (Therneau, 2023) for Cox PH models and base R functions for linear regression. ChatGPT (OpenAI) was used to assist in generating and debugging R code for statistical analyses and data visualization. All code was reviewed, validated, and modified by the authors prior to use.

Results

Descriptive Summary of the Study Population

The primary delayed-entry sample comprised 96 females who survived to age five, of whom 53 (55.2%) died during the observation period and 43 (44.8%) were right-censored at the end of the study period or at last observation. The from-birth sample comprised 176 females followed from birth, of whom 100 (56.8%) died during observation. For AFB analyses, 97 females survived to the 4.5-year biological minimum threshold (80 first births observed, 17 censored) and 64 survived to the six-year threshold (58 first births observed, 6 censored). Regression analyses included between 48 and 96 parous females depending on outcome variable availability and completeness of predictor data. The median age at first birth among parous females was 6.2 years (range: 4.6–11.2 years). Mean reproductive lifespan among females with at least two births was 8.7 years, mean interbirth interval was 2.1 years, and mean offspring survival to age four was 0.61. Across the study population, females experienced a mean of 3.2 alpha male replacements during their first five years of life, 14% experienced group fission before age five, and 8% lost their mothers before age five.

Table 1.1. Descriptive statistics for key life history variables in female white-faced capuchins.

Variable	n	Mean ± SD	Range
Age at First Birth (years)	87	6.4 ± 0.6	5.1–8.2
Reproductive Lifespan (years)	82	5.5 ± 4.6	0.0–16.7
Interbirth Interval (years)	51	575.0 ± 169.7	216.0–960.0
Offspring Survival to Age 4	69	0.56 ± 0.36	0.00–1.00

Note. n = number of individuals with non-missing data for each variable.

Key life history variables for reproductive females at Lomas Barbudal are summarized in Table 1.1. We present descriptive statistics (mean ± SD, range) for all adversity measures in Table 1.2. Continuous variables (group size, mortality rates, rainfall deviation) are summarized with

descriptive statistics (mean, SD, range). For adversity classification, continuous measures were binarized using percentile thresholds (15th/85th for extremes; 85th for high risk), while discrete events (alpha changes, maternal death, fission) are presented as counts and prevalence.

Table 1.2. Descriptive statistics and adversity prevalence across developmental windows.

Window	Variable	N	Mean ± SD	Range	Prevalence (%)
Late Prenatal Window	Rainfall Deviation (z-score)	183	-0.17 ± 0.77	-1.60–3.13	-
	Group Size	174	19.9 ± 7.1	4.0–37.3	-
	Mortality Rate (annualized)	174	0.04 ± 0.14	0.00–0.73	-
	Alpha Male Replacement	184	0.01 ± 0.10	0.00–1.00	-
First Year Window	Rainfall Deviation (z-score)	181	-0.12 ± 0.46	-0.93–1.91	-
	Group Size	165	21.4 ± 6.9	8.0–37.4	-
	Mortality Rate (annualized)	165	0.06 ± 0.18	0.00–2.17	-
	Alpha Male Replacement	184	0.38 ± 1.11	0.00–6.00	-
	Maternal Demise	184	-	-	4.3%
	Fission Event	184	-	-	5.4%
First Five Years Window	Rainfall Deviation (z-score)	181	-0.13 ± 0.37	-0.73–2.86	-
	Group Size	166	20.8 ± 6.1	6.0–36.6	-
	Mortality Rate (annualized)	172	0.06 ± 0.17	0.00–2.17	-
	Alpha Male Replacement (count)	184	2.5 ± 5.5	0–31	-
	Maternal Demise	184	-	-	25.5%
	Fission Event	184	-	-	35.9%

Note. - = not applicable. Rainfall Deviation = absolute deviation z-score (ERA5). Mortality Rate and Group Size are continuous. Maternal demise and fission are reported as prevalence. Alpha-male replacement is reported as specified for each developmental window.

Effects of Early-Life Adversity on Mortality Risk and Age at First Birth

Mortality Risk: Primary Delayed-Entry Models

All three delayed-entry models (Tables S1.5-1.7) had nonsignificant LRT values, indicating that the set of adversity predictors measured during each developmental window did not jointly predict mortality risk among females who survived to age five (Table S1.5). Notably, prenatal alpha male replacement could not be estimated in the delayed-entry model due to complete separation: no female who experienced an alpha replacement during the late prenatal period survived to age five.

In the first-year window model, maternal demise showed a suggestive but nonsignificant association with elevated mortality (HR = 4.62, 95% CI [0.59, 36.25], p = 0.14). In the first-five-year window model, group mortality rate was marginally associated with reduced hazard (HR = 0.67, 95% CI [0.45, 1.00], p = 0.05) but no other ELA variable reached conventional significance thresholds (P<0.05). No delayed-entry model showed a significant global PH violation.

Mortality Risk: From-Birth Sensitivity Analyses

In the prenatal window model (n = 173, 98 deaths; Tables S2–S4), the overall model was not significant by likelihood ratio test (LRT = 4.31 on 4 df, p = 0.37), but alpha male replacement during gestation was a significant individual predictor (HR = 4.33, 95% CI [1.04, 17.95], p = 0.04). The first-year window model (n = 165, 93 deaths) was highly significant overall (LRT = 16.62 on 6 df, p = 0.01), with group mortality rate (HR = 1.42 per SD, 95% CI [1.07, 1.89], p = 0.02) and maternal demise (HR = 4.74, 95% CI [1.84, 12.25], p < 0.01) predicting elevated mortality risk.

The first-five-year window model (n = 166, 93 deaths) was also significant overall (LRT = 14.85 on 6 df, p = 0.02), with group mortality rate predicting elevated mortality risk (HR = 1.84 per SD, 95% CI [1.20, 2.82], p < 0.01). Alpha male replacements showed a marginal association with reduced mortality (HR = 0.96 per replacement, 95% CI [0.93, 1.00], p = 0.06), but no other variables were significant in this developmental period. Global PH tests were nonsignificant for

the prenatal (global $\chi^2 = 8.40$, df = 4, p = 0.08) and first-year (global $\chi^2 = 8.89$, df = 6, p = 0.18) from-birth models (Table S1.4), but significant for the five-year (global $\chi^2 = 19.16$, df = 6, p < 0.01), driven primarily by the mortality-rate ($\chi^2 = 5.50$, p = 0.02) and fission-event ($\chi^2 = 5.89$, p = 0.02) covariates.

Age at First Birth: Primary Models

All three primary AFB models using the 4.5-year biological minimum threshold (Tables S1.8-S1.10) had p-values above 0.05 with a likelihood ratio test. No individual predictor approached significance in any model. PH tests were non-significant for the prenatal **(global p = 0.41) and first-year (global p = 0.42) models; the five-year model showed a global violation (p = 0.03**), driven by the effect of group size **(χ^2=6.58, df=1,** p = 0.**0**1) (Table S1.10).

Age at First Birth: Six-Year Sensitivity Analyses

In the six-year threshold sensitivity analyses (Tables S1.11-S1.13), significant predictor associations were limited to the first-five-year developmental window, for which the overall model was marginally significant (LRT p = 0.05; Table S1.11). Each additional alpha replacement experienced during the first five years of life was associated with an ~ 7% higher hazard of first birth (HR = 1.07 per replacement, 95% CI [1.02, 1.12], p < 0.01). Maternal demise likewise predicted accelerated reproduction (HR = 2.34, 95% CI [1.13, 4.87], p < 0.01). Group fission showed a marginal trend toward delayed reproduction (HR = 0.60, 95% CI [0.33, 1.10], p = 0.10). Rainfall deviation, group size, and group mortality rate did not predict AFB in any six-year threshold model, regardless of developmental window.

Life History Regressions

The results of these analyses are presented in Figure 2. Only one of twelve models reached overall significance (Table S1.14-1.16, in footnote): the first-year model predicting offspring survival to age four ($F(6,61) = 2.46$, $P = 0.03$, $R^2 = 0.19$, adjusted $R^2 = 0.12$. No associations survived the more conservative window-wide FDR correction for any of our 12 models.

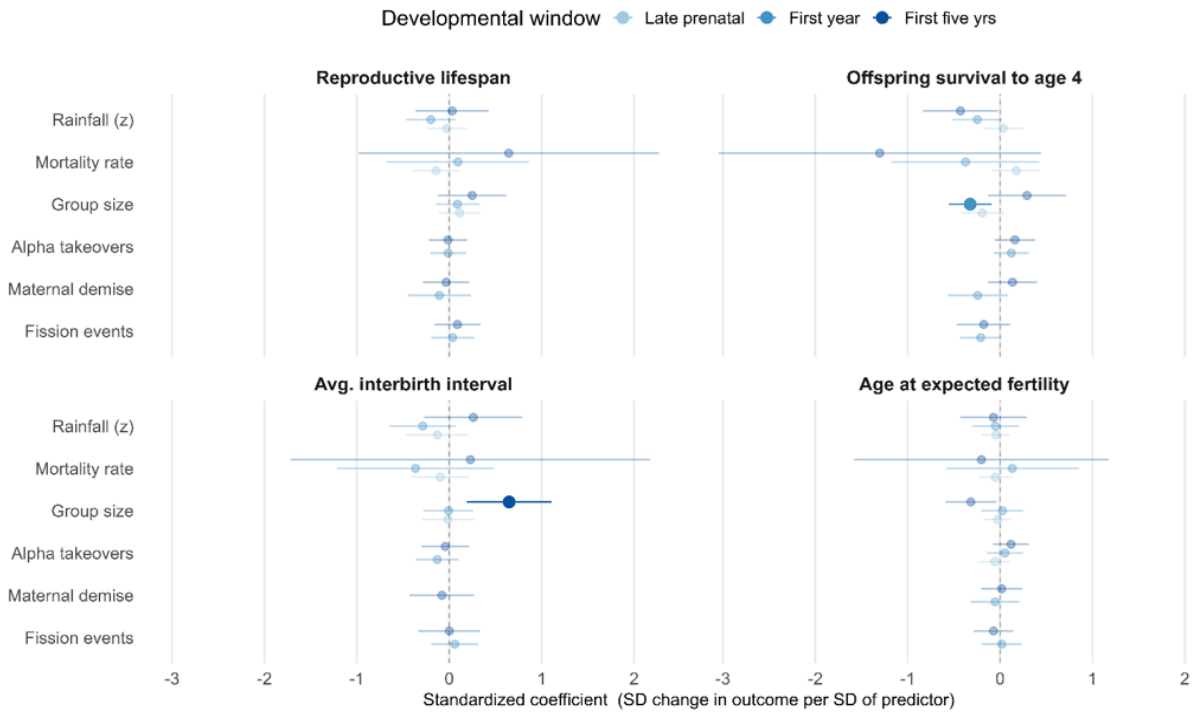


Figure 1.1. Early-life predictors of female life-history outcomes. Standardized regression coefficients and 95% confidence intervals are shown for associations between early-life rainfall, mortality rate, group size, alpha male takeovers, maternal demise, and group fission events and four female life-history outcomes: reproductive lifespan, offspring survival to age 4, average interbirth interval, and age at expected fertility. Estimates are shown separately for exposures occurring during late prenatal development, the first year of life, and the first five years of life. Coefficients are standardized as the standard deviation (SD) change in the outcome per SD increase in the predictor. Enlarged solid points indicate associations significant at a false discovery rate (FDR)-adjusted $p<0.05$ within predictor.

Following within-predictor FDR correction, group size during the first year (Table S1.15) predicted lower offspring survival to age four ($\beta = -0.12$, SE = 0.04, $P = 0.03$). In the first-five-year window (Table S1.16), larger group size was associated with longer average interbirth intervals ($\beta = 110.02$, SE = 39.54, $P = 0.03$) and lower age-expected fertility ($\beta = -0.37$, SE = 0.16, $P = 0.05$), although the overall models for both outcomes were nonsignificant. Rainfall during the first five years was negatively associated with offspring survival before multiple-

testing correction ($\beta = -0.15$, SE = 0.07, $p = 0.04$), but this association did not survive within-predictor FDR correction (P = 0.17). Prenatal models yielded no significant associations after FDR correction (Table S1.14).

Discussion

The present study identifies three recurring patterns. First, primarily ecological adversities (extreme rainfall values, large group size, and high group mortality rates) and primarily psychosocial adversities (alpha male replacement, maternal demise, fission) were generally associated with different adult reproductive life-history outcomes. Second, these associations differed markedly between from-birth and delayed-entry analyses, indicating that conditioning on survival substantially alters the patterns observed among adults. Third, psychosocial adversity**,** but not the ecological exposures measured here**,** was associated with accelerated age at first birth in the age-six sensitivity analysis. Together, these patterns are consistent with distinct developmental pathways involving developmental constraint, stress-mediated reproductive acceleration, and viability selection, although the underlying physiological mechanisms remain to be tested directly.

Comparison of Delayed-Entry and From-Birth Models

The comparison between from-birth and delayed-entry model specifications reveals a pattern that is both methodologically critical and biologically informative: every significant predictor in the from-birth mortality models becomes nonsignificant or inestimable in the corresponding delayed-entry model, which conditions on survival to age five. Prenatal alpha male replacement cannot be estimated in the delayed-entry model at all, due to complete separation - no female who experienced an alpha replacement during the late prenatal period survived to age five. The only difference between these model specifications is whether individuals who died before

age five are included or excluded, and the dramatic attenuation of effect sizes demonstrates that these early-dying individuals are precisely those driving the adversity–mortality associations in the full sample.

This pattern reflects the fact that ELA exposures that are sufficiently severe to kill remove the most affected individuals from the population during or immediately after the developmental window, before any lasting "programming" effects on later-life outcomes could manifest. Critically, viability selection and developmental programming are mutually exclusive pathways for any given individual - if an exposure kills an individual during the developmental period, there is no post-exposure period for programming to operate. The from-birth models capture the *total* effect of ELA, lethality plus any programming that occurs among survivors, while the delayed-entry models capture only the *programming* effect among individuals who survived the exposure period. The difference between these two estimates is the lethality component of the exposure itself.

Early-Life Adversity Is Often Lethal

From-birth survival models revealed substantial mortality associated with prenatal alpha male replacement, first-year maternal loss, and first-five-year group mortality rate. Such juvenile mortality represents a strong selective filter: individuals lacking the capacity to survive these adversities are removed from the reproductive pool before they can contribute to future generations. However, for ELA to shape the *evolution* of adaptive developmental plasticity rather than merely acting as a demographic filter, the exposure must generate differential fitness among survivors in ways that are predictable from early-life conditions. The contrast with delayed-entry models reflects this filtering. Associations with first-year group mortality rate and maternal demise were attenuated or became nonsignificant after conditioning on survival to age five, while prenatal

alpha male replacement could not be estimated because no exposed female survived to that age. Group mortality rate in the first-five-year window is a notable exception to this pattern: rather than becoming nonsignificant, it remains significant in the delayed-entry model but reverses direction. However, differences in standardization across risk sets make this reversal somewhat difficult to interpret. The large effect sizes observed here suggest that these ELA exposures impose substantial selection, but the evolutionary response depends on whether the plastic adjustments that buffer against early mortality also carry fitness consequences (e.g., in reproductive timing or allocation) that manifest in adulthood.

This pattern reflects how viability selection can obscure developmental associations among survivors. Adversity severe enough to kill individuals during or shortly after the developmental window removes them from the population before later-life phenotypes can be expressed (Morrison et al., 2023). Viability selection and developmental programming can therefore occur in response to the same exposure, but mortality determines which programmed phenotypes remain observable at later ages. Consequently, estimates derived from survivor-only samples more broadly cannot be interpreted independently of the viability selection that determined which individuals remained available to express programmed phenotypes. Viability selection and developmental plasticity therefore need not represent alternative explanations for the consequences of ELA. Environmental hazards that impose differential mortality simultaneously create the selective conditions under which mechanisms of developmental programming evolve. Plasticity is therefore expressed only among individuals for whom physiological adjustment is sufficient to maintain viability. Developmental programming and viability selection may therefore operate sequentially across the life course, rather than as competing explanations for the consequences of early adversity.

Psychosocial Adversity and the Timing of First Reproduction

Among females surviving to age six, alpha male replacements and maternal loss during the first five years were associated with earlier first reproduction, whereas the ecological exposures measured here were not. Importantly, these associations emerged only in the six-year sensitivity analysis and were absent from the primary 4.5-year models (which began follow-up at the earliest possible age of reproduction, when many females have not yet entered the population's typical reproductive period). These findings are consistent with a broader literature linking instability in the early social environment to variation in reproductive development, including associations of father absence, family conflict, and social threat with earlier reproductive development in humans (Webster et al., 2014). Alpha male replacements should not, however, be treated as a non-human analogue of father absence as they represent a change in the social regime of the group, but do not necessarily entail paternal absence.

Alpha male replacements are better characterized as episodes of social instability that may involve intense male-male conflict, changes in dominance and affiliative relationships, altered maternal behavior, and elevated infanticide risk. Repeated replacements during development may consequently expose females to recurrent changes in the structure and predictability of their social environment. Maternal demise represents a different but similarly profound disruption of the early social environment, involving the permanent loss of a central social partner, as well as, at younger ages, potential energetic and caregiving consequences. The fact that these two qualitatively different psychosocial exposures were associated with reproductive timing raises the possibility that instability in the developmental social environment is relevant to later reproductive scheduling. Repeated exposure to social instability may alter the developmental physiology or

behavioral maturation in ways that become consequential once females reach reproductive maturity, although the present study cannot identify such a mechanism directly.

Furthermore, we note that maternal demise is itself a confounded event - mothers who die during a daughter's early development may also have been in poorer condition, embedded in more unstable social networks, or exposed to correlated ecological stressors prior to death, any of which could independently affect an offspring's reproductive timing. Our analyses cannot distinguish the consequences of maternal loss itself from these correlated maternal and environmental conditions. These estimates should be read as suggestive rather than definitive, as the overall 6-year threshold age at first birth sensitivity model for the five year developmental window is only marginally significant (LRT = 12.41, p-value = 0.05).

Ecological Adversity and Reproductive Output

We acknowledge the concern that psychosocial and ecological adversity categories are not perfectly orthogonal. Social instability may affect foraging efficiency through displacement from preferred feeding sites, competitive exclusion from food patches, and disruption of coordinated group movement, while resource scarcity may exacerbate social conflict through increased contest competition for limited food items. The correlation between psychosocial and ecological adversity measures in our data is modest but nonzero, reflecting these biological connections. Nevertheless, the differing patterns across outcomes support the analytical usefulness of distinguishing these exposure classes, while not implying that each measure operates cleanly through only one causal pathway.

Group mortality rate predicts juvenile mortality across developmental windows, while psychosocial adversity (alpha replacements, maternal loss) carries the largest individual effects. Rainfall deviation and group size did not independently predict mortality risk in any model

specification, suggesting that their constraining effects on life history operate through pathways other than direct lethality - possibly through energetic limitation that reduces reproductive capacity without elevating mortality. The constraint pattern is internally consistent: ecological adversity imposes energetic costs during development that may limit the somatic resources available for lifetime reproduction, resulting in fewer years of reproductive activity and lower survival of the offspring that are produced. Psychosocial adversity (alpha replacements, maternal loss) also kills during early life but additionally accelerates reproductive timing among survivors, advancing age at first birth without affecting reproductive lifespan or offspring survival. This acceleration is consistent with the possibility that early-life stress alters later reproductive timing through HPA–HPG axis cross-talk and developmental programming of the neuroendocrine systems regulating puberty (Camille Melón & Maguire, 2016; Livadas & Chrousos, 2019). This divergence in downstream consequences provides empirical justification for distinguishing adversity types, even if the developmental exposures themselves are correlated. The empirical results help resolve the theoretical concern about category overlap by demonstrating that the distinction has predictive validity in that they produce different life history signatures.

Developmental Constraint and Predictive Calibration Operate Simultaneously

These findings resolve, at least for this population, the apparent contradiction between acceleration and constraint models of ELA effects on life history. Our results demonstrate that both processes operate, but through different adversity types and with different downstream consequences. Ecological adversity imposes developmental constraints, energetic deprivation during development reduces the somatic resources available for lifetime reproduction, leading to shorter reproductive lifespans and lower offspring survival among parous females.

Psychosocial adversity, by contrast, both kills (removing the most vulnerable before reproductive age) and accelerates reproduction among survivors. This dual-process framework, constraint through ecological adversity, stress-mediated reproductive programming through psychosocial adversity, reconciles findings that have appeared contradictory when adversity types are lumped into composite indices. Studies using cumulative adversity scores may find predominant constraint effects because their indices weight ecological and psychosocial components equally, and the constraint signal from ecological adversity may dominate the weaker calibration signal from psychosocial adversity. Studies focusing specifically on social adversity (e.g., maternal loss, social instability) may find acceleration effects because they isolate the adversity type through which calibration operates. The apparent inconsistency in the literature may thus reflect methodological differences in adversity measurement rather than genuine biological inconsistency across species or populations.

The first-five-year window emerges as the critical integrative period for both processes, likely because it reflects a period during which exposure to the specific ELA variables measured here are less lethal. The relative weakness of prenatal and first-year effects on adult life history (except for the strong first-year mortality effects) suggests that brief developmental windows, while sufficient to impose acute mortality risk, may be too short to durably program the reproductive system-or, alternatively, that the from-birth models for these windows are dominated by the lethality signal, with any surviving-programmer signal too weak to detect in the present sample. Viability selection adds a layer of complexity to this picture by removing the most adversity-affected individuals before they reach adulthood. Given that adversity determines impacts survival, the null results in our delayed-entry mortality and primary AFB models are a predictable outcome of conditioning on a fixed survival threshold.

Limitations

Several limitations warrant careful consideration. Complete separation for prenatal alpha replacement in the delayed-entry model reflects the fact that no female exposed to an alpha replacement during the late prenatal period survived to age five. This precludes estimation of any residual programming effect among survivors. Alpha male replacement captures heterogeneous social disruptions that may include infanticide of dependent offspring, coalitionary aggression among adult and subadult males, changes in group ranging patterns and habitat use, disruption of established affiliative relationships, and altered maternal behavior in response to the new male's presence. We cannot isolate which components of the replacement process drive the observed associations with mortality and reproductive timing.

Furthermore, sample sizes for regression analyses and the six-year AFB sensitivity analysis are modest, though they are comparable to samples in other long-term primate field studies. While the effects detected in these analyses are statistically robust after FDR correction, smaller effects may go undetected, and some estimated have wide confidence intervals that reflect the imprecision of estimates from moderate samples. Continued data collection at Lomas Barbudal would improve statistical power for these analyses in future years as additional females enter the parous sample.

Finally, our rainfall data are derived from the ERA5 reanalysis dataset rather than from on-site weather station measurements for the full study period. While ERA5 data have been validated against on-site measurements during overlapping periods and show adequate correspondence (Table S1), gridded climate products may smooth local variation in precipitation, potentially attenuating the true association between rainfall conditions experienced at the study site and developmental outcomes. Future analyses incorporating higher-resolution local climate data,

should they become available for the full study period, may yield even stronger rainfall effects than those reported here.

Conclusions

Early-life adversity in wild female capuchin monkeys operates through multiple concurrent pathways whose divergent effects may be obscured by approaches that promote the use of a cumulative ELA measure. Among survivors, ecological adversity constrains lifetime reproductive output through lasting energetic limitation, while psychosocial adversity accelerates reproductive timing under conditions of social instability. These results demonstrate that constraint and stress-mediated reproductive acceleration are not competing explanations for ELA effects on life history; both operate simultaneously through different adversity types, different developmental mechanisms, and different life history outcomes.

Supplementary Information

This Supplementary Information provides detailed operational definitions for all life-history and early-life adversity variables used in the manuscript, documentation of rainfall validation procedures, supplementary statistical analyses, proportional hazards diagnostics, mixed-sex sensitivity analyses, and full regression outputs.

Female Life History Variables

Reproductive Lifespan

Reproductive lifespan is determined as the difference in years between a female's first birth and their last birth. Estimation of these ages is described in more detail below:

Age at First Birth

Age at first birth is determined by identifying the earliest verified birth for each female capuchin monkey while accounting for potential observation gaps that could obscure reproductive onset. We begin by excluding females who experienced >60-day monitoring gaps during their prime first-birth window (ages 4.5–7 years – calculated based on records of earliest and latest known 'first births' within this population) *and* who lack any recorded offspring born after those gaps. This ensures that only females with reliable reproductive histories are included. For eligible females, the mother's age at first reproduction is computed by comparing her offspring's earliest possible birth date to her own birth date, using only females for whom birth dates are known within a six-month window of certainty. This approach minimizes false negatives from poor monitoring while capturing biologically meaningful variation in reproductive timing.

Age at Last Birth

Age at last birth is determined by identifying the most recent verified birth for each deceased female capuchin monkey while accounting for potential gaps in observation. We begin by linking mothers to their last known offspring using birth records, then calculate the mother's age (in years) at that birth by comparing the offspring's birth date to the mother's birth date. To ensure reliability, we exclude cases where the mother's death occurred after an observation gap exceeding 60 days unless she had a surviving infant under 1 year old at the time of the gap (a circumstance in which we can be sure it is very unlikely to have missed a potential birth due to the existence of an unweaned infant). The final output is the mother's age at her last confirmed reproduction, adjusted for uncertainty in birth/death dates and observation gaps.

Average Interbirth Interval

Average interbirth interval is calculated by first identifying all consecutive pairs of births for each mother in the population. Each pair consists of a previous birth and subsequent birth. To ensure accuracy, intervals that overlap with periods where observations may have been missed are excluded from analysis. These excluded periods are defined as gaps longer than 60 days where the mother was unobserved.

It is not always sufficient to assume no birth occurred during a gap by the absence of an infant when observations resume due to the possibility that said infant does not survive until the end of the gap. It is unlikely that females who went unobserved for a period of less than 60 days gave birth unexpectedly during that period due to the obvious physical changes that accompany the last 60 days of pregnancy. However, not all observation gaps greater than 60 days are likely to contain a missed birth. We may have a high degree of certainty that a gap greater than 60 days does *not* contain a birth if the gap is entirely contained within one year of the previous birth and

the infant from the previous birth is still alive (interbirth intervals <1 year are only observed in cases where the previous infant is lost). We may also expect that a gap greater than 60 days does not contain a birth if the gap is entirely contained within the 'non-birthing season' which occurs between September to December. If a longer gap occurs between February and August but is entirely contained within 150 days (a conservative estimate of the late prenatal period, typically 162 days) prior to and immediately following an observed birth, we can similarly be highly certain that a birth was not missed. Only intervals that occur during well-observed periods are retained for further analysis.

For mothers with valid birth intervals remaining after filtering, we calculate two versions of the average interbirth interval to account for uncertainty in birth dates. The conservative estimate uses the latest possible date for the previous birth and earliest possible date for the subsequent birth, resulting in the shortest plausible average interval. The liberal estimate does the opposite, using the earliest possible date for the previous birth and latest possible date for the subsequent birth to calculate the longest plausible average. The difference between these two estimates provides a quantitative measure of uncertainty. We mark the beginning and end of each birth sequence to avoid double-counting intervals that might overlap across different birth pairs. The lengths of the remaining non-discarded interbirth intervals were then averaged within female. The final output includes mother-specific average interbirth intervals that reflect only periods of reliable observation, along with measures of uncertainty.

Age Expected Fertility

Age-expected fertility is the predicted number of births for a female capuchin monkey at a given age, calculated by applying population-level age-specific fertility rates to her observed

exposure time (adjusted for observation gaps), allowing us to assess whether an individual's reproductive output aligns with or deviates from the group average for their age.

For this analysis we use *birthing seasons,* rather than calendar age, because capuchin monkeys exhibit seasonal breeding patterns, with births concentrated in specific months. A female's reproductive success is more biologically tied to her position within this annual cycle than to her exact calendar age: for example, a female who turns 6 years old just before the birthing season has nearly a full season of reproductive potential at that age, whereas one who turns 6 just after the season must wait a year. To align with this biology, we define a female's "age" based on birthing seasons anchored to her own birth date: each season runs from October 15 to the following October 15 (based on observed peaks in conception). This ensures all females are assessed at equivalent points in their reproductive cycles, avoiding artifacts caused by arbitrary calendar-year boundaries. We use the word 'age' to simplify description, however it should be noted that for this variable we are discussing age in terms of birthing seasons, rather than calendar-years.

To calculate *population-average births per age*, we aggregate data across all females. First, each female's contribution to a given age is weighted by her *adjusted exposure time* (time observed minus observations gaps). Births are assigned to specific ages based on their overlap with the female's birthing seasons, with fractional credit for uncertain dates. The total births at age X (summing all females' births) are then divided by the total female-years of exposure at age X (summing all females' adjusted exposure days ÷ 365.24). This yields an age-specific fertility rate (births/female/year) that accounts for both seasonal timing and observation gaps, providing a population-level estimate of reproductive patterns across ages.

This metric discounts periods where births could have been missed. For example, a female with 10 years of exposure but 2 years of gaps would only contribute 8 years of adjusted exposure to her expected fertility calculation. This adjustment ensures the estimate reflects real-world observation constraints, making it a more reliable benchmark for comparing individual reproductive success. By contrasting this age-expected fertility with observed births, we are able to identify females whose fertility deviates from population norms.

Adjusting for Observation Gaps

To account for periods when births could have gone undetected, we systematically adjust exposure time by identifying observation gaps (e.g., due to field season breaks/temporary absences). Gaps in observation are flagged using census data, and we record the start and end dates of intervals where undetected births might have occurred (e.g., between a female's last non-pregnant sighting and her next appearance with an infant, or any period >150 days). For each female, the total exposure time during a given age-specific birth season is reduced by the number of days overlapping with these gaps.

It is common that the exact date of birth may be missed by observers. In these cases, we estimate a range of birth dates based on the first possible day a birth may have occurred and the last possible date (in both cases accounting for the estimated age of the infant when first observed and the length of the preceding observation gap). Births with uncertain dates that occur during an observation gap and span possible birth seasons are fractionally allocated to adjacent seasons based on the proportion of the birth date range falling outside the gap. For instance, if 60% of the possible birth date range falls within the birthing season a female at age 6 and 40% in age 7, then 0.6 births are counted for age 6 and 0.4 for age 7.

This yields an estimate of expected births adjusted by exposure, which predicts fertility rates using only reliably observed time, preventing artificial inflation or underestimation of reproductive success for females with patchy data.

Total Number of Births

The *minimum and maximum total number of births* estimates a female's total reproductive output by combining confirmed births with projections of potentially missed births during observation gaps. For deceased females, the minimum count is derived directly from observed offspring birth records, that is records of known live births. The maximum count adds estimated missed births during gaps (e.g., between reproductive maturity and first observed birth, or between consecutive births), calculated by dividing gap durations by the population's average interbirth interval (593–699 days). For example, a 10-year gap could accommodate 5–6 additional births. We also record a range of uncertainty over the total number of births as the difference between minimum observed and maximum number of births (the minimum observed plus the maximum total number of missed births). In doing so, we explicitly separate verified data (observed births) from projections (gap-filled estimates), ensuring conservative bounds for lifetime fertility. Only deceased females are included to ensure complete life histories.

Proportion of Offspring Surviving to Age 4

The proportion of offspring surviving to age 4 is calculated by identifying all offspring born to mothers with precisely known birth seasons. For each offspring, survival status is determined by comparing birth and death dates (or last observation dates for living individuals), accounting for date uncertainty: offspring are classified as *confirmed survivors* if their minimum possible age exceeds 4 years, *confirmed deaths* if their maximum possible age is under 4 years, or *uncertain* if their age range spans 4 years (only 4 cases). Living offspring without death dates

are assigned a cutoff year (2030) to calculate maximum age. We aggregate offspring counts per mother (total live births, confirmed survivors, and confirmed deaths) and exclude mothers in groups with incomplete monitoring. The final proportion is derived as the proportion of offspring known to have lived to age 4 to the total number of live births, with uncertainty bounds for rare ambiguous cases.

End dates for each social group: Not all social groups were studied continuously from 1990-2026. Each group had a separate end date beyond which visits to the group were considered too rare for the demographic data to be reliable; a female's end date for her demographic outcome variables (i.e. the date at which we declare her to be right censored in survival analyses) corresponds to the last date of reliable censusing for the group in which she was last a member.

Early Life Adversity

Average Group Size

Monkeys are counted as members of a group if they have been recorded in that group over five consecutive censuses on days where the group was observed for at least 6 hours. Their date of membership within a group is set as the date of the first of those five consecutive censuses. Similarly, a monkey is considered to have 'left' a group (either via emigration or death) once their absence has been recorded over five consecutive censuses on days with at least 6 hours of observation. In such instances, their date of departure/death is set to the day immediately following the first of those five consecutive 'absence' censuses. Each monkey associated with a group is then assigned an interval during which they can reliably be counted as a member of that group and are assumed to have been present in that group during all days within that interval, even if no census has been taken and the group went unobserved.

Group assignments were manually checked against birth records for individuals with a low number of census observations and the following adjustments were made: a) if the monkey is considered to have entered the group prior to its first birthday, and the group 'entered' according to census records is its natal group, then the date of entry into the group for that individual is changed to their estimated birth date; b) if a monkey is never observed in 5 consecutive censuses within any group, but were seen within their natal group at an age <1 year, they are counted as members of that group from their estimated birth date until the day immediately following their last day seen within that group. For each monkey, we calculate the average size of their social group during three critical developmental windows: (1) the 50-day prenatal period (using the mother's group membership), (2) the first year of life, and (3) the first five years of life. To calculate average group size during early-life for each individual, we sum the number of 'monkey/days' over the interval for which the individual is recorded as a member of the group. "Monkey/days" are the number of monkeys within a group on any given day within the interval during which the individual is considered a member of that group. For example, if there are 3 monkeys in a group every day across the interval 2000-01-01 to 2001-01-05 this would represent 15 'monkey/days'. This number is then divided by the number of days the individual is recorded as a member of that same group (so for this example, 15 / 5 = 3). Thus 'average group size' is calculated over each early-life interval of interest (birth – 50 days, first year of life, first five years of life). For the last-trimester prenatal period, mother's group membership is used to produce an estimate of average group size.

We made one manual adjustment to the individual 'monkey/days' during the prenatal period for monkey 'MO341' whose mother ('FB0063') dies shortly after MO341's birth. We can be confident that 'MO341' was born to 'FB0063' within the recorded natal group, however due

to a lack of censuses with observation hours >6 during this critical period and the timing of'FB0063''s death, the total days spent by the mother within MO341's natal group during the prenatal period is likely to be underestimated. We used census records on days with <6 observation hours to confirm the mother's presence in the natal group prior to MO341's birth and manually set the total individual 'monkey/days' for MO341 to '50' over this interval. This method allows us to most accurately account for the comings and goings of other group members, including births, deaths, and emigrations and to account for possible group membership changes that the individual may experience (due to fission events) which may occur during the interval of interest. However, it may exclude temporary migrants or individuals who die very young and thus are not observed in a study group for at least 5 consecutive censuses. In both of these cases, neither are likely to represent significant sources of intragroup feeding competition over the early-life intervals of interest and thus their exclusion is not a measurement concern.

Group Mortality Rate

Monkeys are counted as members of a group if they have been recorded in that group over five consecutive censuses on days where the group was observed for at least 6 hours, following the same criteria used for calculating average group size. Their date of membership begins on the first of those five consecutive censuses, and their departure (via emigration or death) is recorded after five consecutive absences.

However, for mortality calculations two modifications are made to the process for calculating average group size described above: 1) Infants who die young are included even if they were recorded in fewer than five censuses before death and 2) Adult males are excluded from both the mortality count and the group exposure time. As males are the dispersing sex, it is

often difficult, if not impossible, to determine whether their absence is due to emigration or death, therefore we excluded all males who were at least 5 years of age (an average estimate for early emigrants) across each interval from both the calculation of 'monkey-days' and deaths. This conservative approach avoids misclassifying male dispersal events as deaths, which could artificially inflate mortality estimates.

For each individual monkey over a given interval (e.g., the first five years of life), we calculate the mortality rate they were exposed to within their group by first determining the total group monkey-days, which represents the cumulative exposure time of all group members during the focal monkey's residency. For every day the focal monkey was a confirmed group member, we sum the number of other monkeys present in the group (excluding males >5 years old). For example, if a monkey was in a group for 5 days with 3 other members each day, the total group monkey-days would be 15. Next, we count the number of confirmed deaths (excluding adult males) among these group members during the same interval, including infants who died before being recorded in five consecutive censuses. The resulting mortality rates are calculated as the ratio of deaths to total group monkey-days, producing a daily mortality rate that can be annualized by multiplying by 365. This yields an intuitive measure expressed as deaths per monkey-year of exposure. For example, a value of 0.5 indicates one death would be expected for every two monkey-years of observed group exposure.

This approach accounts for temporal variation in group size, differences in observation periods across individuals, and the aggregated death rate experienced by the group during the focal monkey's residency. By weighting exposure time and deaths consistently, it provides a robust measure of the mortality pressure each monkey experienced during early-life interval, -

mirroring the logic of average group size calculations but applied to mortality risk rather than social density.

Alpha Male Replacements

Alpha male replacements, or takeovers, are detected by cross-referencing the date of alpha male transitions with individual monkeys' group residency. Unlike other rank reversals, which are more subtle, the change in the alpha position is generally dramatic, both in the amount of social conflict leading up to the change and in the morphological and behavioral changes occurring in the new and recently deposed alpha males. The new alpha male engages in more displays (breaking branches, ostentatious urine rubbing, gratuitous aggression especially against other species), becomes more socially central, and is constantly piloerect so that his body size seems to increase; the former alpha male instantly deflates in apparent size by ceasing piloerection and retreats from the center of the group. The timing of these changes is noted in the census data. For each monkey, we assess whether takeovers occurred in their social group during three critical developmental windows: (1) the 50-day prenatal period (using the mother's group membership), (2) the first year of life, and (3) the first five years of life.

To ensure accuracy, a takeover is only counted if its date falls within both the target interval (ex: first year of life) *and* a period when the monkey (or its mother, for the prenatal window) was confirmed to reside in the group-verified by census-based residency records. We generate two complementary metrics: a binary indicator (any takeover during the target interval: yes/no) and a count of total takeovers per interval. For the prenatal window, we use the mother's group residency to approximate the infant's prenatal environment. Age at first takeover is calculated by comparing the earliest postnatal takeover date to the monkey's birth date, adjusted for uncertainty in birth timing.

This method accounts for group transfers (e.g., fission events or early dispersal events) by tracking all groups a monkey inhabited before age 5. It excludes takeovers that occurred during gaps in residency or in groups the monkey never actually occupied. The result is a time-anchored measure of social instability experienced by each monkey during early life.

Maternal Demise

This variable identifies whether an individual's mother died during their first year or first five years of life, while also estimating the offspring's age at maternal loss. For each monkey in the study, we link their birth records to their mother's death record (if it exists) and calculate: (1) the precise age of maternal loss (using best estimate birth and death dates), (2) conservative age boundaries (using the earliest and latest birth and death date estimates to account for uncertainty), and (3) binary indicators for maternal demise during the first year or five years of the individual's life. If the mother's death window (earliest to latest possible date) overlaps with the offspring's developmental window (e.g., first year after birth), the event is counted. This ensures we capture all plausible cases while handling date uncertainty. For example, if a mother died between 11–13 months after her offspring's birth, the offspring would be flagged for first-year maternal demise, as the earliest possible death date (11 months) falls within this window.

Fission Events

The date of fission was determined by examining long-term census data and behavioral observations. We identified the fission event as the permanent division of one social group into two distinct groups that are no longer regularly associated. The process often involved a period of instability with back-and-forth movement of individuals between the emerging groups. The fission date was assigned as the last day the group was observed intact, immediately followed by the first observation of the groups as separate without subsequent reunification. In cases of

prolonged instability, we used the date after which the groups were never observed together again. When there were extended observation gaps, we used the midpoint of the gap or the first observation of the split if the gap was too long to determine the exact date.

Climatic Perturbations

Climate data were extracted from ERA5, a global atmospheric reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECMWF; https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5). Reanalysis products combine historical observations from multiple sources with a physically based numerical weather model to generate spatially and temporally complete, consistent estimates of past climatic conditions. ERA5 provides estimates of precipitation on an approximately 31-km global grid and is widely used in climate, environmental, hydrological, and ecological research (Hersbach et al., 2020; Stefanidis et al., 2021; Huang et al., 2025). ERA5 provides complete monthly precipitation estimates spanning the full study period without gaps, making it preferable to local weather station data, which are incomplete across the study period. We validated ERA5 against available on-site weather station records from the Lomas Barbudal region during the overlapping period (September 1996 – September 2015; n = 229 months) and found strong agreement (Pearson $r = 0.83$, Spearman $\rho = 0.87$, both $p < 0.001$; mean bias = +18mm). Although ERA5 shows a small systematic positive bias relative to station measurements, this does not affect our analyses, as all rainfall predictors are expressed as z-score deviations from long-term means. Monthly precipitation values were used for all analyses. To align monthly averages with developmental periods of interest, the following steps were taken: (1) for the last 50 days of gestation, we included all months overlapping this window (e.g., if the last 50 days of gestation spanned March 15–May 5, we used data from March-May); (2) for postnatal periods, monthly sums were aligned

with birth months (e.g., a monkey born June 10 would have Year 1 calculated from June 01 to June 30 of the following year, but using whole-month aggregates); (3) when developmental windows did not align with calendar months (e.g. a 50-day prenatal period spanning parts of 3 months), we employed a weighted averaging using day-based fractional scaling approach (this procedure is described in full below).

Climatic perturbations were quantified as deviations from historical rainfall patterns during the developmental periods previously described. For each individual, monthly precipitation totals derived from the ERA5 climate dataset were converted to z-scores relative to the 1979-1989 baseline period (mean ± SD for each calendar month). This baseline period was chosen because it occurs prior to the study start (1990) and includes an interval across which we have both local weather station data and ERA5 estimates. Three exposure metrics were calculated by averaging these z-scores across each developmental window: (a) late prenatal (50 days pre-birth), (b) first year (birth to 365 days), and (c) years 0-5 (0-1,825 days). The z-score transformation (observed minus historical mean, divided by historical SD) standardizes comparisons across years, where positive values indicate wetter-than-average conditions and negative values indicate drought. While real numbers for each z-score transformation were maintained for post-hoc analyses, we use the absolute value of z-scores to facilitate analysis, using a percentile rank to identify individuals who experienced particularly extreme perturbations during development compared to what might be expected for this population. For individuals with uncertain birth/death dates (recorded as ranges), conservative estimates were derived using the latest plausible birth date and earliest plausible death date.

Weighted Z-Scores for Partial Months During Prenatal Period

Since the 50-day prenatal period capturing the 'last trimester' (from 50 days pre-birth to birth date) often spans parts of two or three calendar months, our analysis weighted partial months that overlap with this window. For example, a monkey born June 15 would have a prenatal window from April 26–June 15, which partially covers April (last 5 days), all of May, and the first 15 days of June. To construct a weighted z-score, we scaled each month's contribution by the fraction of days actually falling within the prenatal window (e.g., April's climate data would be weighted at 5/30 of its full monthly value), ensuring periods with greater biological relevance have proportionally greater influence on the final exposure metric while maintaining complete temporal coverage of the 50-day window. The following equation summarizes our approach:

$$Weighted\ Z = \frac{\Sigma \frac{n}{i=1} \left(Z_i \times \frac{Days_i^{in\ window}}{Days_i^{in\ month}}\right)}{\Sigma \frac{n}{i=1} \frac{Days_i^{in\ window}}{Days_i^{in\ month}}}$$

To illustrate the impact of this approach over an unweighted z-score, consider a monkey born June 15, 2005, for whom the following exposure data applies:

Month	**Days in Window**	**Days in Month**	**Raw Z-score**	**Weighted Contribution**
Apr	5 (Apr 26-30)	30	+2.1	2.1 x (5/30) = 0.35
May	31	31	-1.3	-1.3 x (31/31) = -1.3
Jun	15 (Jun 1-15)	30	+0.7	0.7 x (7/30) = 0.35

In this example, we would calculate the weighted Z-score for the last trimester as follows:

$$Weighted\ Z = \frac{0.35 + (-1.3) + 0.35}{\frac{5}{30} + \frac{31}{31} + \frac{15}{30}} = \frac{-0.6}{1.67} = -0.36$$

We can then compare this to unweighted average Z-score for the same individual:

$$\frac{2.1 + (-1.3) + 0.7}{3} = +0.5$$

The weighted result better captures May's dominant influence and 'downweighs' edge months (in this case, April).

Figure S1.1: Predictor Variable Correlations by Developmental Period

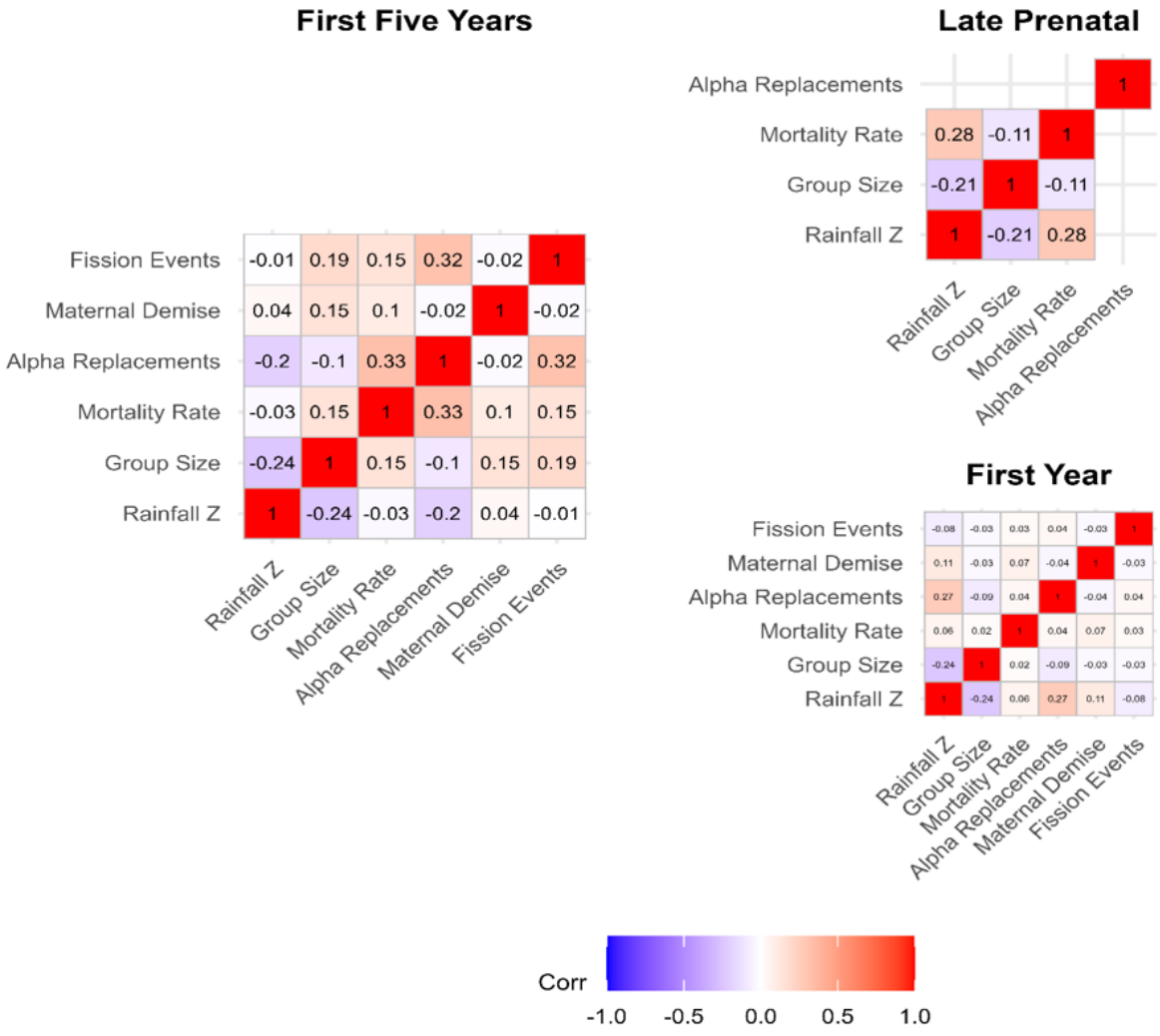


Figure S1.1. Correlations among predictor variables by developmental period. Pairwise Pearson correlation coefficients among early-life predictor variables are shown for the late prenatal period, first year of life, and first five years of life. Values within cells indicate Pearson's correlation coefficient (r), with color indicating the direction and magnitude of the association. Rainfall was expressed as a standardized z-score. Variance inflation facrors (VIFs) were caluclaed saparately for each developmental-period model; all VIFs were <2.

Validation of ERA5 Precipitation Data Against On-Site Weather Station Records

To verify that ERA5 accurately represents precipitation conditions at the study site, we compared ERA5 monthly totals against available weather station records from the Lomas Barbudal region during periods of overlap.

Data sources

ERA5 monthly precipitation totals were extracted for the grid cell containing the Lomas Barbudal study site (10.50°N, 85.37°W), spanning January 1990 through December 2023. Weather station records were obtained from three stations in the region: the on-site station at the Lomas Barbudal Biological Reserve (10.506°N, 85.371°W) and the Palo Verde station (10.383°N, 85.334°W), located approximately 14 km south of the study site, and the official weather station for Liberia located at the Daniel Obuder Quirós International Airport (10.354°N, 85.324°W), approximately 21 km northwest of the study site. The Lomas station is the most directly relevant to our study site but has limited temporal coverage; the Palo Verde station provides a substantially longer record and serves as the primary validation comparison. Validation was restricted to months with available station data.

Validation procedure

ERA5 monthly totals were matched to station records by calendar month and year, retaining only months with non-missing values in both sources. Agreement was assessed using: (1) Pearson correlation (r), with 95% confidence intervals; (2) Spearman rank correlation (ρ); (3) mean bias (ERA5 − station, mm); (4) mean absolute error (MAE, mm); and (5) root mean squared error (RMSE, mm). Bland-Altman analysis was used to visualize systematic bias and limits of agreement across the range of precipitation values. All analyses were conducted in R version 4.3.1.

Results

ERA5 was strongly correlated with station records at both locations (Table S1.1; Figures S1.2-1.4). At the primary Palo Verde station (September 1996 – September 2015; n = 229 months), ERA5 yielded Pearson r = 0.83 (95% CI: 0.78, 0.86) and Spearman ρ = 0.87 ($p < 0.001$), with a small positive bias relative to station measurements (mean bias = +18mm; MAE = 52mm; RMSE = 83mm). At the on-site Lomas station (n = 29 months), agreement was equally strong (Pearson r = 0.83, 95% CI: 0.66, 0.92; Spearman ρ = 0.89; $p < 0.001$), with a small negative bias (mean bias = −19mm; MAE = 57mm; RMSE = 105mm). Finally, the Liberia station data (January 1980 – December 2010; n = 252 months) evidenced a similarly strong pattern of agreement with the ERA5 precipitation estimates (Pearson r = 0.85, 95% CI: 0.82, 0.88; Spearman ρ = 0.93; $p < 0.001$), with a small negative bias (mean bias = −19.81 mm; MAE = 49.88mm; RMSE = 94.87mm). The modest bias observed at all stations is consistent with the known tendency of reanalysis products to smooth extreme precipitation events in tropical regions. As all rainfall predictors in this chapter are expressed as z-score deviations from long-term means, this additive bias does not affect the standardized measures used in analyses.

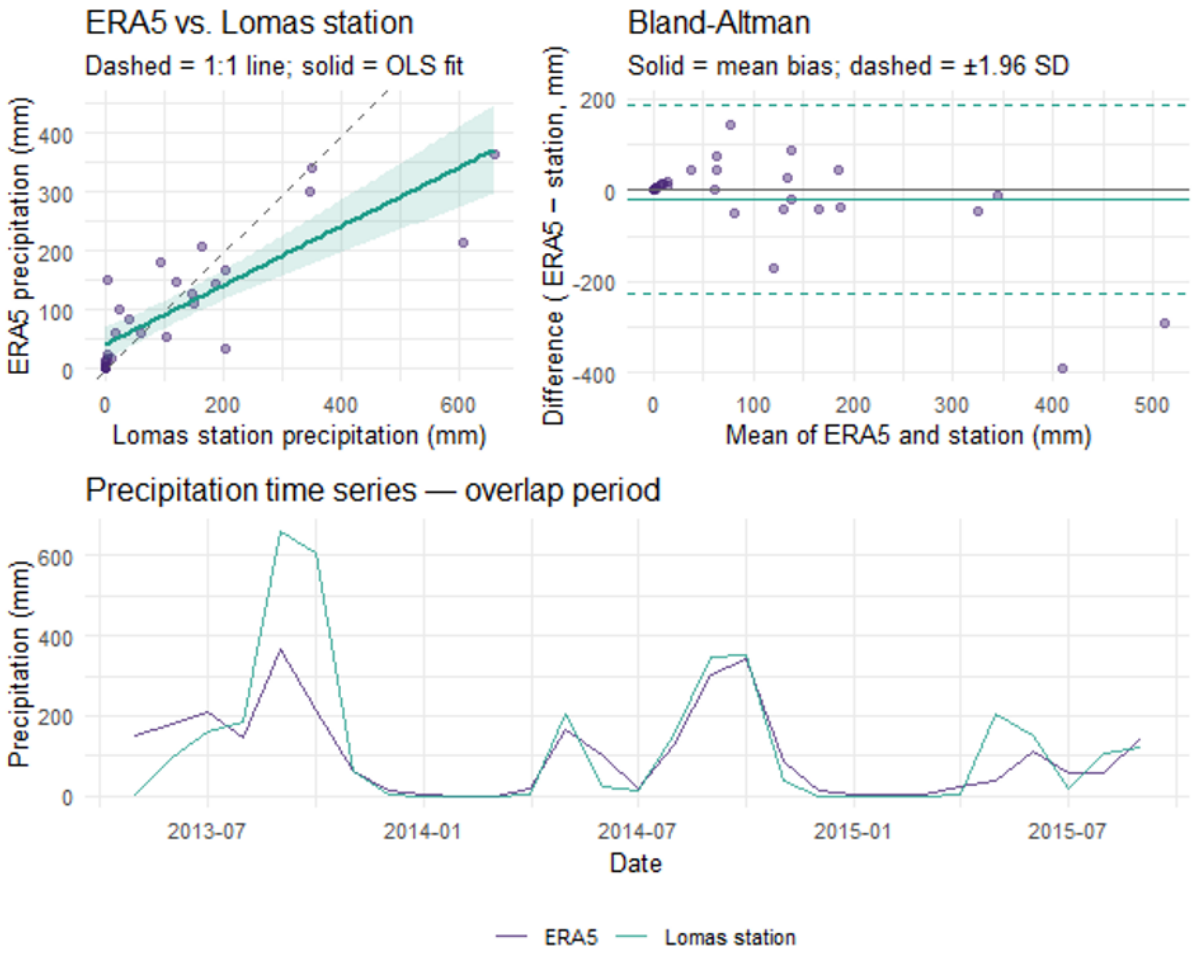


Figure S1.2. Validation of ERA5 monthly precipitation estimates against weather station observations from Lomas Barbudal Biological Reserve. The upper left panel compares monthly precipitation totals derived from ERA5 reanalysis data with values recorded at the Lomas Barbudal weather station during the period of overlap. The dashed line indicates the 1:1 relationship, while the solid line represents the OLS regression fit with 95% confidence intervals shaded. The upper right panel presents a Bland–Altman analysis of agreement between ERA5 and station measurements, with the solid horizontal line indicating mean bias and dashed lines representing ±1.96 standard deviations of the differences. The lower panel shows the monthly precipitation time series for both datasets across the overlap period. ERA5 generally tracked the temporal dynamics of local rainfall but underestimated several high-precipitation months relative to station observations.

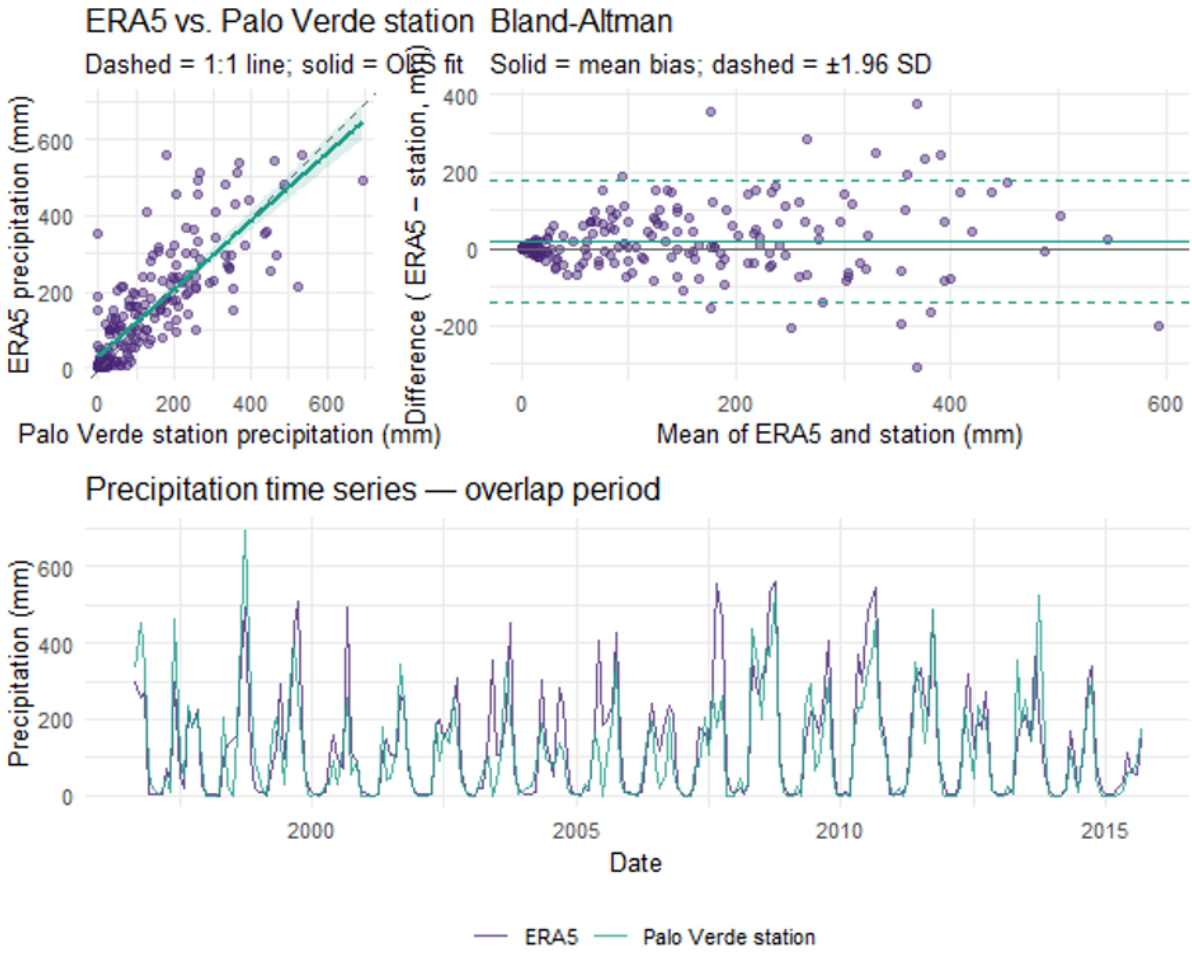


Figure S1.3. Validation of ERA5 monthly precipitation estimates against weather station observations from Palo Verde National Park. The upper left panel compares monthly precipitation totals derived from ERA5 reanalysis data with values recorded at the Palo Verde weather station during the period of overlap. The dashed line indicates the 1:1 relationship, while the solid line represents the ordinary least squares regression fit with 95% confidence intervals shaded. The upper right panel presents a Bland–Altman analysis of agreement between ERA5 and station measurements, with the solid horizontal line indicating mean bias and dashed lines representing ±1.96 standard deviations of the differences. The lower panel shows the monthly precipitation time series for both datasets across the overlap period.

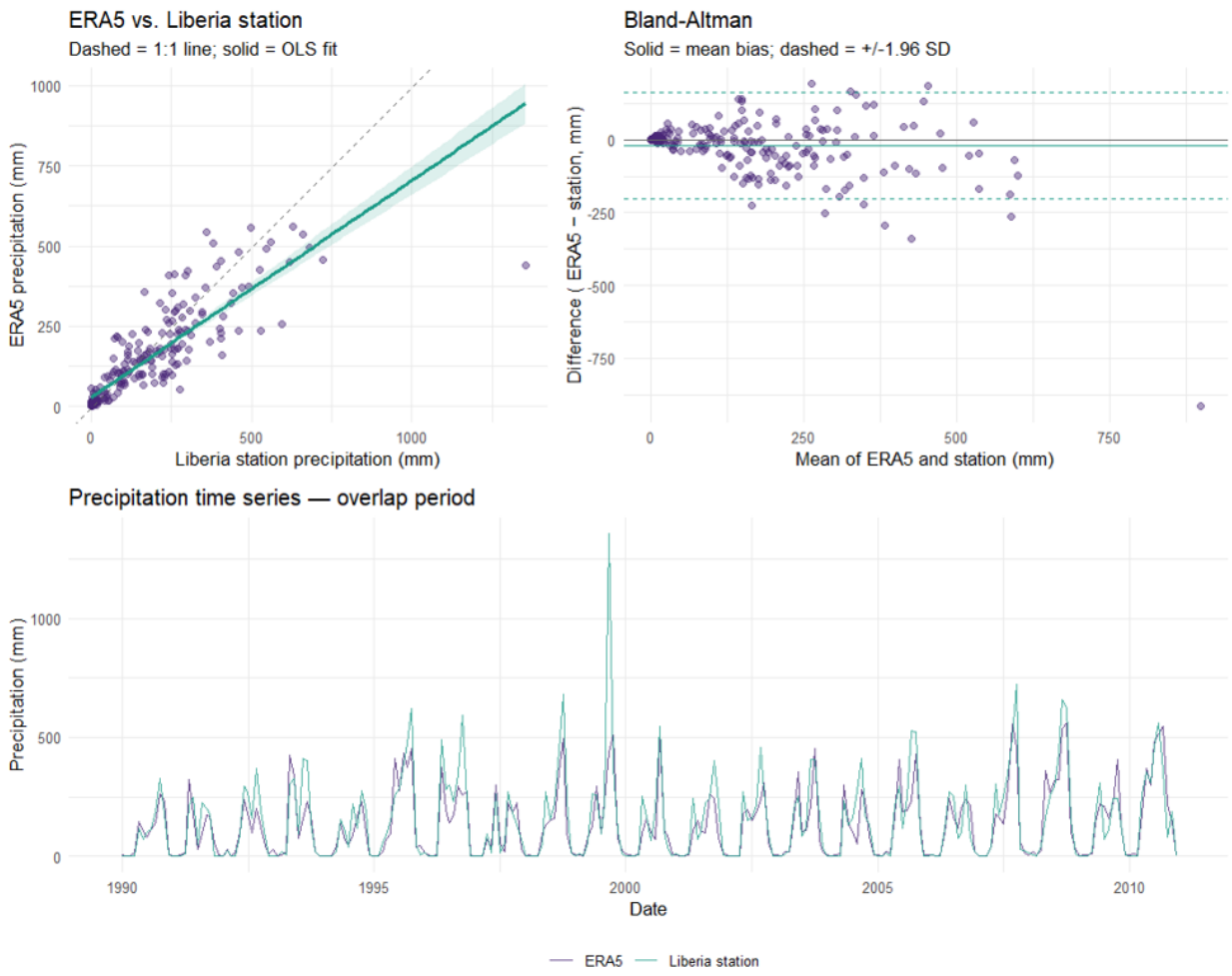


Figure S1.4. Validation of ERA5 monthly precipitation estimates against weather station observations from Liberia weather station. The upper left panel compares monthly precipitation totals derived from ERA5 reanalysis data with values recorded at the Liberia weather station during the period of overlap from 1980-01-01 to 2010-12-01. The dashed line indicates the 1:1 relationship, while the solid line represents the ordinary least squares regression fit with 95% confidence intervals shaded. The upper right panel presents a Bland–Altman analysis of agreement between ERA5 and station measurements, with the solid horizontal line indicating mean bias and dashed lines representing ±1.96 standard deviations of the differences. The lower panel shows the monthly precipitation time series for both datasets across the overlap period.

Table S1.1. Validation of ERA5 monthly precipitation against weather station records.

Station	Pearson r (95% CI)	Spearman p	OLS slope	Mean bias (mm)	Mean Adjusted Error (mm)	Root Mean Standard Error (mm)	N
Lomas Barbudal	0.828 (0.662, 0.916)	0.889	0.501	-19.04	57.04	105.11	29
Palo Verde	0.826 (0.780, 0.864)	0.871	0.889	17.78	51.74	83.08	229
Liberia Historic	0.857 (0.820, 0.887)	0.936	0.676	-19.81	49.88	94.87	252

Note: OLS slope reflects agreements with a 1:1 relationship, not just correlation strength. A slope well below 1 indicates ERA5 underestimates the station's largest events even when r is high.

Table S1.2. Cox Proportional Hazards From-Birth Mortality Risk Model Fit Statistics

Model	n	Deaths	Likelihood Ratio Test	df	P-Value
Late Prenatal	173	98	4.31	4	0.37
First Year	165	93	16.62	6	**0.01**
First Five Years	166	93	14.85	6	**0.02**

Note. From-birth models (no left-truncation), females only. Sample sizes reflect complete cases on each developmental window's covariate set. Bold P-values indicate $p < 0.05$.

Table S1.3. From-Birth Mortality Models - Hazard Ratios by Developmental Window

Developmental Window	Predictor	HR	95% CI	P-Value	Sig.
Late Prenatal	Rainfall deviation (z-score)	0.86	0.64–1.14	0.30	
	Group Size	1.00	1.00–1.00	0.96	
	Mortality Rate	1.10	0.91–1.33	0.33	
	Alpha Male Replacements	**4.33**	**1.04–17.95**	**0.04**	*
First Year	Rainfall deviation (z-score)	1.47	0.88–2.46	0.14	
	Group Size	1.00	1.00–1.00	0.56	
	Mortality Rate	**1.42**	**1.07–1.89**	**0.02**	*
	Alpha Male Replacements	1.01	0.87–1.18	0.88	
	Maternal Demise	**4.74**	**1.84–12.25**	**< 0.01**	**
	Fission Event	1.37	0.65–2.89	0.41	
First Five Years	Rainfall deviation (z-score)	0.51	0.19–1.40	0.19	
	Group Size	1.00	0.99–1.00	0.55	
	Mortality Rate	**1.84**	**1.20–2.82**	**< 0.01**	**
	Alpha Male Replacements	0.96	0.93–1.00	0.06	.
	Maternal Demise	1.28	0.78–2.12	0.33	
	Fission Event	0.97	0.59–1.61	0.91	

Note. HR = Hazard Ratio. 95% CI = 95% confidence interval. Bold values indicate $p < 0.05$. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.10$. Mortality rate and group size are rescaled within each developmental-window sample. Rainfall is the signed weighted deviation (positive = wetter, negative = drier).

Table S1.4. Results of Proportional Hazards Assumption Tests for From-Birth Mortality Models

Model	Variable	χ^2	df	P-Value
Late Prenatal	Rainfall deviation (z-score)	0.15	1	0.70
	Group Size	**4.45**	**1**	**0.04**
	Mortality Rate	0.14	1	0.71
	Alpha Male Replacements	2.94	1	0.09
	Global Test	8.40	4	0.08
First Year	Rainfall deviation (z-score)	0.17	1	0.68
	Group Size	3.05	1	0.08
	Mortality Rate	2.15	1	0.14
	Alpha Male Replacements	0.07	1	0.79
	Maternal Demise	2.43	1	0.12
	Fission Event	0.06	1	0.80
	Global Test	8.89	6	0.18
First Five Years	Rainfall deviation (z-score)	0.90	1	0.34
	Group Size	0.01	1	0.93
	Mortality Rate	**5.50**	**1**	**0.02**
	Alpha Male Replacements	1.44	1	0.23
	Maternal Demise	2.59	1	0.11
	Fission Event	**5.89**	**1**	**0.02**
	Global Test	**19.16**	**6**	**< 0.01**

Note. Tests are based on scaled Schoenfeld residuals. Bold rows indicate $p < 0.05$.

Table S1.5. Cox Proportional Hazards Delayed-Entry Mortality Risk Model Fit Statistics

Model	n	Deaths	Likelihood Ratio Test	df	P-Value
Late Prenatal	93	51	3.04	3	0.39
First Year	96	53	6.41	6	0.38
First Five Years	96	53	7.07	6	0.31

Model	n	Deaths	Likelihood Ratio Test	df	P-Value

Note. Delayed-entry Cox models with left-truncation at age 5. Females only. Bold P-values indicate $p < 0.05$.

Table S1.6. Delayed-Entry Mortality Models - Hazard Ratios by Developmental Window

Developmental Window	Predictor	HR	95% CI	P-Value
Late Prenatal	Rainfall deviation (z-score)	0.97	0.66–1.42	0.86
	Group Size	1.00	1.00–1.01	0.08
	Mortality Rate	1.03	0.79–1.34	0.82
	Alpha Male Replacements	–	–	complete separation
First Year	Rainfall deviation (z-score)	1.70	0.84–3.47	0.14
	Group Size	1.00	1.00–1.01	0.08
	Mortality Rate	1.01	0.80–1.29	0.91
	Alpha Male Replacements	1.07	0.89–1.28	0.48
	Maternal Demise	4.62	0.59–36.25	0.14
	Fission Event	1.35	0.53–3.46	0.53
First Five Years	Rainfall deviation (z-score)	3.14	0.86–11.39	0.08
	Group Size	1.00	0.99–1.01	0.81
	Mortality Rate	0.67	0.45–1.00	**0.05**
	Alpha Male Replacements	1.02	0.97–1.07	0.48
	Maternal Demise	0.72	0.31–1.65	0.44
	Fission Event	1.03	0.51–2.08	0.93

Note. HR = Hazard Ratio. 95% CI = 95% confidence interval. Bold values indicate $p < 0.05$. Rainfall operationalized as signed ERA5 deviation. – = predictor not estimable due to complete separation (Alpha Male Replacement, Late Prenatal Window).

Table S1.7. Results of Proportional Hazards Assumption Tests for Delayed-Entry Mortality Models

Model	Variable	χ^2	df	P-Value
Late Prenatal	Rainfall deviation (z-score)	0.45	1	0.50
	Group Size	1.49	1	0.22
	Mortality Rate	0.76	1	0.38
	Global Test	3.34	3	0.34
First Year	Rainfall deviation (z-score)	0.01	1	0.93
	Group Size	1.02	1	0.31
	Mortality Rate	0.00	1	0.97
	Alpha Male Replacements	0.30	1	0.58
	Maternal Demise	2.59	1	0.11
	Fission Event	0.22	1	0.64
	Global Test	4.37	6	0.63
First Five Years	Rainfall deviation (z-score)	0.47	1	0.49
	Group Size	0.11	1	0.74
	Mortality Rate	0.88	1	0.35
	Alpha Male Replacements	2.26	1	0.13
	Maternal Demise	1.06	1	0.30
	Fission Event	1.27	1	0.26
	Global Test	10.58	6	0.10

Note. Tests are based on scaled Schoenfeld residuals. Bold rows indicate $p < 0.05$.

Table S1.8. Age at First Birth - Model Fit Statistics (4.5-year threshold)

Model	N	# events	Likelihood Ratio Test	df	P-Value
Late Prenatal	94	78	1.50	3	0.68
First Year	97	80	2.88	6	0.82
First Five Years	97	80	2.75	6	0.84

Note. Primary models use a 4.5-year threshold. Bold P-values indicate $p < 0.05$.

Table S1.9. Age at First Birth - Hazard Ratios (4.5-year threshold)

Developmental Window	Predictor	Hazard Ratio	95% CI	p-value
Late Prenatal	Rainfall Z-score	0.91	0.71–1.17	0.47
	Group Size	1.00	1.00–1.00	0.98
	Mortality Rate	0.94	0.73–1.20	0.61
First Year	Rainfall Z-score	1.00	0.59–1.70	1.00
	Group Size	1.00	1.00–1.01	0.37
	Mortality Rate	1.01	0.79–1.29	0.97
	Alpha Male Replacements	0.92	0.78–1.09	0.35
	Maternal Demise	1.81	0.24–13.55	0.57
	Fission Events	0.79	0.31–2.01	0.62
First Five Years	Rainfall Z-score	1.25	0.50–3.14	0.63
	Group Size	1.00	0.99–1.01	0.87
	Mortality Rate	0.95	0.76–1.19	0.68
	Alpha Male Replacements	1.03	0.98–1.07	0.22
	Maternal Demise	1.37	0.75–2.48	0.30
	Fission Events	0.83	0.51–1.37	0.47

Note. HR = Hazard Ratio. 95% CI = 95% confidence interval. Bold p-values indicate $p < 0.05$.

Table S1.10. Tests of the Proportional Hazards Assumption for Primary Age at First Birth Models (4.5-year threshold)

Developmental Window	Predictor	χ^2	df	p-value
Late Prenatal	Rainfall Z-score	1.92	1	0.17
	Group Size	0.42	1	0.52
	Mortality Rate	0.47	1	0.49
	Global Test	2.89	3	0.41
First Year	Rainfall Z-score	1.58	1	0.21
	Group Size	0.13	1	0.72
	Mortality Rate	0.08	1	0.78
	Alpha Male Replacements	1.92	1	0.17
	Maternal Demise	2.40	1	0.12
	Fission Events	0.61	1	0.44
	Global Test	6.03	6	0.42
First Five Years	Rainfall Z-score	0.90	1	0.34
	Group Size	**6.58**	**1**	**0.01**
	Mortality Rate	0.89	1	0.35
	Alpha Male Replacements	3.40	1	0.07
	Maternal Demise	1.32	1	0.25
	Fission Events	1.62	1	0.20
	Global Test	**13.63**	**6**	**0.03**

Note. Tests are based on scaled Schoenfeld residuals. Bold rows indicate $p < 0.05$.

Table S1.11. Age at First Birth Sensitivity Model Fit Statistics (6-year threshold)

Model	N	# events	Likelihood Ratio Test	df	P-Value
Late Prenatal	62	57	0.48	3	0.92
First Year	64	58	4.93	6	0.55
First Five Years	64	58	12.41	6	0.05

Note. Sensitivity models use a 6-year threshold. Bold P-values indicate $p < 0.05$.

Table S1.12. Age at First Birth - Hazard Ratios (6-year threshold)

Developmental Window	Predictor	Hazard Ratio	95% CI	p-value
Late Prenatal	Rainfall Z-score	0.98	0.75–1.28	0.873
	Group Size	1.00	1.00–1.00	0.857
	Mortality Rate	0.94	0.71–1.24	0.647
First Year	Rainfall Z-score	1.08	0.58–2.01	0.812
	Group Size	1.00	1.00–1.01	0.284
	Mortality Rate	0.87	0.63–1.20	0.411
	Alpha Male Replacements	0.92	0.76–1.12	0.421
	Maternal Demise	4.20	0.53–33.28	0.175
	Fission Events	0.62	0.19–2.07	0.442
First Five Years	Rainfall Z-score	1.49	0.52–4.27	0.461
	Group Size	1.00	0.98–1.01	0.383
	Mortality Rate	0.95	0.72–1.26	0.736
	Alpha Male Replacements	1.07	1.02–1.12	**0.007****
	Maternal Demise	2.34	1.13–4.87	**0.022***
	Fission Events	0.60	0.33–1.10	0.099.

Note. HR = Hazard Ratio. 95% CI = 95% confidence interval. Bold values indicate $p < 0.05$. Significance codes: ** $p < 0.01$, * $p < 0.05$, . $p < 0.10$.

Table S1.13. Tests of the Proportional Hazards Assumption for Sensitivity Age at First Birth Models (6-year threshold)

Developmental Window	Predictor	χ^2	df	p-value
Late Prenatal	Rainfall Z-score	0.47	1	0.50
	Group Size	0.70	1	0.40
	Mortality Rate	0.12	1	0.73
	Global Test	1.61	3	0.66
First Year	Rainfall Z-score	1.43	1	0.23
	Group Size	0.00	1	0.98
	Mortality Rate	3.78	1	0.05
	Alpha Male Replacements	2.37	1	0.12
	Maternal Demise	2.55	1	0.11
	Fission Events	0.00	1	0.96
	Global Test	9.30	6	0.16
First Five Years	Rainfall Z-score	0.08	1	0.78
	Group Size	2.40	1	0.12
	Mortality Rate	0.13	1	0.71
	Alpha Male Replacements	0.14	1	0.71
	Maternal Demise	0.01	1	0.92
	Fission Events	0.01	1	0.92
	Global Test	3.47	6	0.75

Note. Tests are based on scaled Schoenfeld residuals. Bold rows indicate $p < 0.05$.

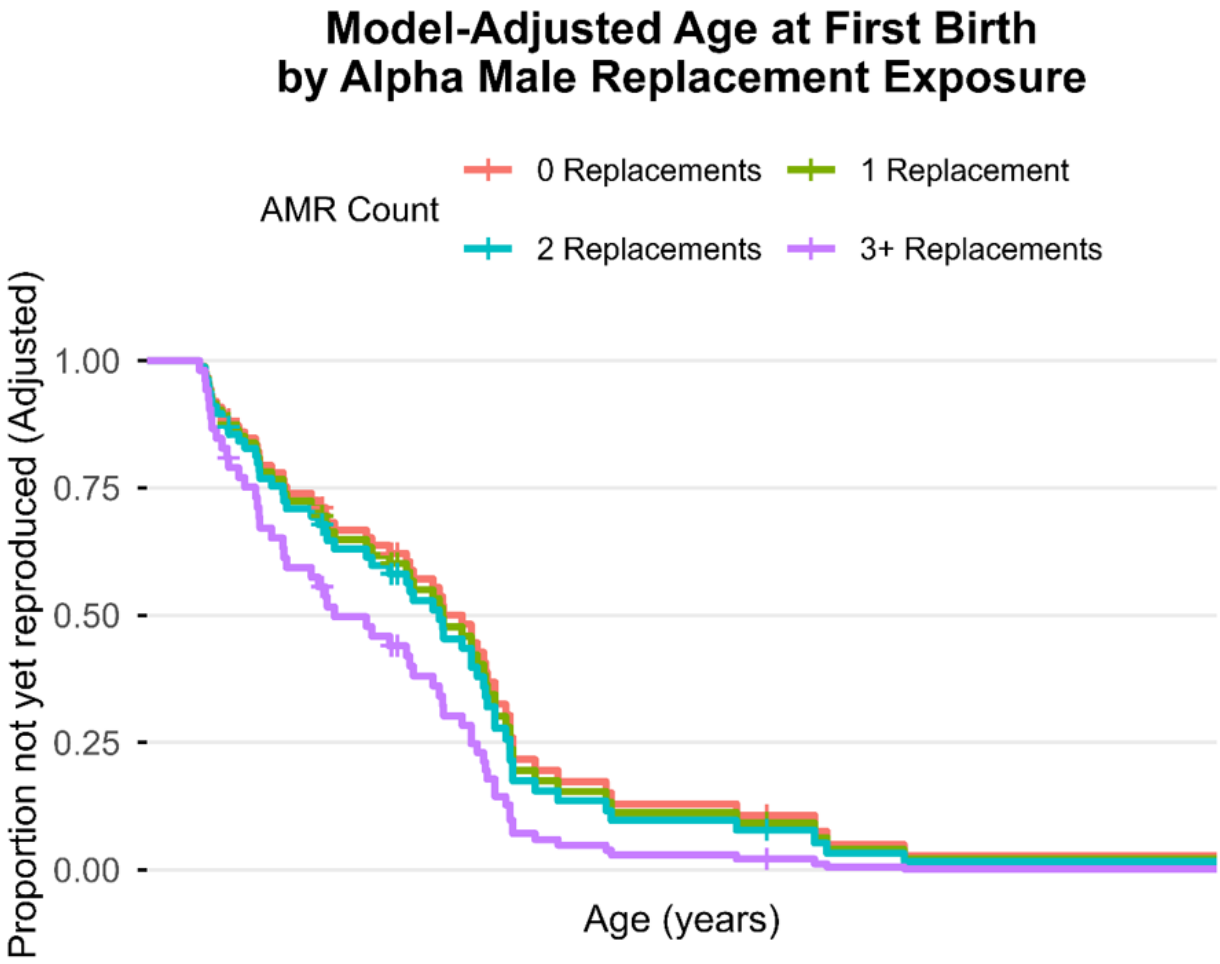


Figure S1.5Model-adjusted age at first birth by alpha male replacement exposure. Model-adjusted survival curves show the estimated proportion of females not yet reproduced as a function of age according to exposure to alpha male replacements during the first five years of life. Curves are shown for females exposed to 0, 1, 2, and 3 or more alpha male replacements. For the 3+ category, predictions were generated using the observed mean number of replacements among females exposed to at least three replacements. Other model covariates were held at their sample means, except maternal demise and group fission, which were set to 0. Curves were estimated from the six-year sensitivity Cox proportional hazards model..

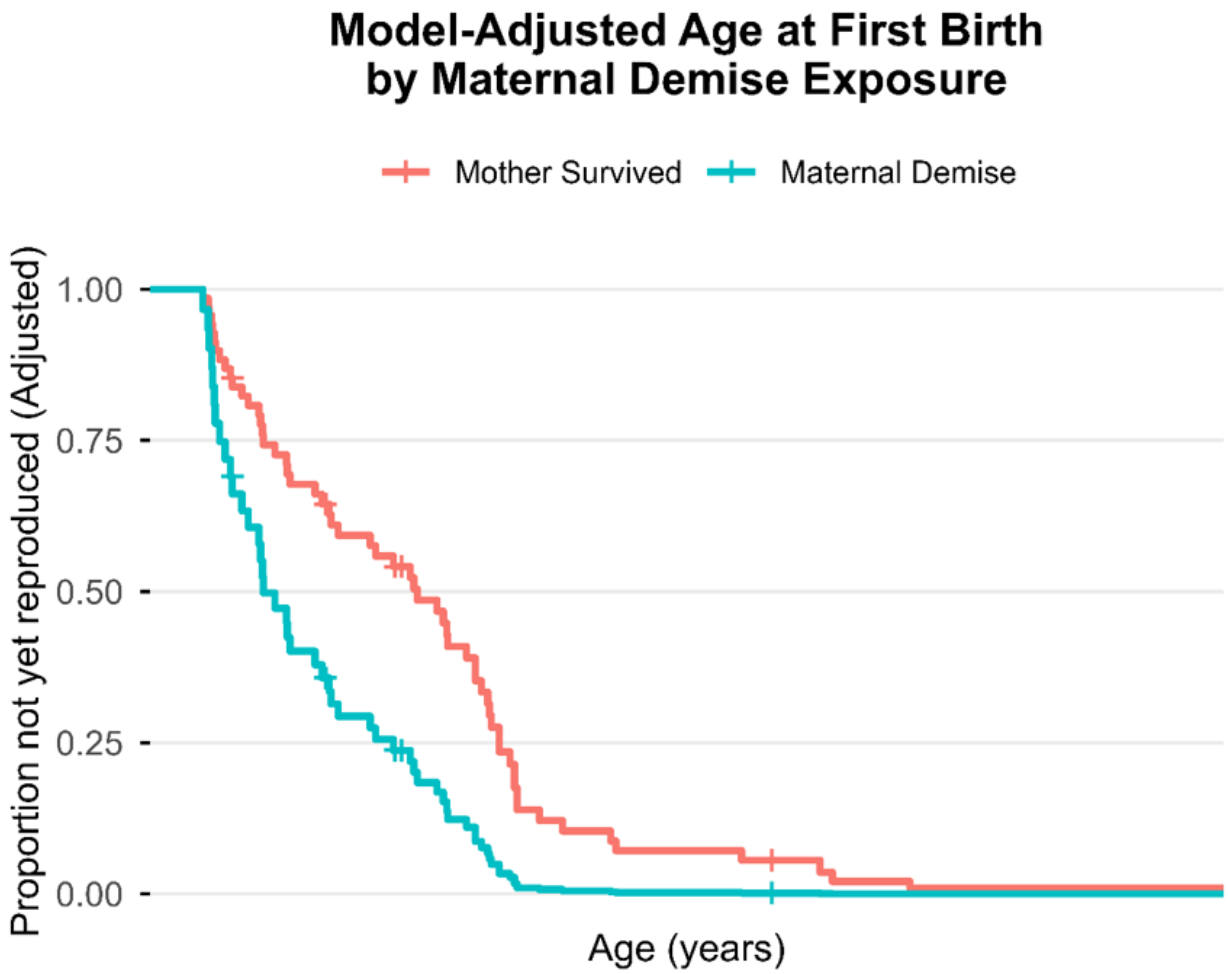


Figure S1.6. Model-adjusted age at first birth by maternal demise exposure. Model-adjusted survival curves show the estimated proportion of females not yet reproduced as a function of age according to exposure to whether their mother survived or died during the female's first five years of life. Other model covariates were held at their sample means, except maternal demise and group fission, which were set to 0. Curves were estimated from the six-year sensitivity Cox proportional hazards model.

Drought Exposure and Early Infant Mortality: Fisher's Exact Test and From-Birth Cox Models

To characterize the drought-mediated early mortality bottleneck that produces complete drought selection in the delayed-entry sample, we examined the association between first-year rainfall drought exposure and mortality before age one using the full mixed-sex population sample. Drought exposure was defined as a negative first-year ERA5 z-score relative to the 1979–1989 baseline.

Table S1.14. Drought prevalence by early mortality status

	Drought-exposed (first-year z < 0)	Not drought-exposed	Total

Died before age 1	61	76	137
Survived or censored after age 1	40	348	388
Total	101	424	525

Fisher's Exact Test: OR = 6.94, 95% CI: 4.25 – 11.49, $p < 0.001$. Individuals dying before age one had nearly seven times higher odds of having experienced below-average first-year rainfall than those who survived or were censored after age one.

Table S1.15. Late Prenatal Period Predictors of Life History Outcomes

Predictor	Outcome	N	Estimate	Std. Error	t	p (raw)	p (FDR within-predictor)	p (FDR window-wide)	Sig.
Rainfall (signed z)	Reproductive Lifespan	79	-0.13	0.51	-0.25	0.80	0.80	0.87	
Group Size	Reproductive Lifespan	79	0.51	0.53	0.96	0.34	0.68	0.84	
Mortality Rate	Reproductive Lifespan	79	-0.65	0.60	-1.09	0.28	0.56	0.84	
Alpha Takeovers	Reproductive Lifespan	*79*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	
Rainfall (signed z)	Average Interbirth Interval	48	-19.71	29.21	-0.67	0.50	0.80	0.84	
Group Size	Average Interbirth Interval	48	-0.81	24.30	-0.03	0.97	0.97	0.97	
Mortality Rate	Average Interbirth Interval	48	-15.88	27.00	-0.59	0.56	0.58	0.84	
Alpha Takeovers	Average Interbirth Interval	*48*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	
Rainfall (signed z)	Offspring Survival to Age 4	66	0.01	0.04	0.32	0.75	0.80	0.87	
Group Size	Offspring Survival to Age 4	66	-0.07	0.04	-1.59	0.12	0.47	0.84	
Mortality Rate	Offspring Survival to Age 4	66	0.06	0.05	1.33	0.19	0.56	0.84	
Alpha Takeovers	Offspring Survival to Age 4	*66*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	
Rainfall (signed z)	Age-Expected Fertility	93	-0.05	0.09	-0.60	0.55	0.80	0.84	
Group Size	Age-Expected Fertility	93	-0.03	0.09	-0.38	0.70	0.94	0.87	
Mortality Rate	Age-Expected Fertility	93	-0.06	0.11	-0.56	0.58	0.58	0.84	
Alpha Takeovers	Age-Expected Fertility	93	-0.56	0.84	-0.67	0.51	0.51	0.84	

Note. Linear regression results testing associations between late-prenatal predictors and adult life history outcomes. Rainfall is coded as the signed deviation z-score (ERA5), in which positive values indicate wetter-than-average and negative values drier-than-average conditions. Alpha takeovers were not estimable in this window. Model fit: Reproductive Lifespan, $R^2 = 0.04$, adj. $R^2 = -0.00$, $F(3,75) = 1.00$, $p = 0.40$; Average Interbirth Interval, $R^2 = 0.03$, adj. $R^2 = -0.04$, $F(3,44) = 0.43$, $p = 0.73$; Offspring Survival to Age 4, $R^2 = 0.09$, adj. $R^2 = 0.04$, $F(3,62) = 1.97$, $p = 0.13$; Age-Expected Fertility, $R^2 = 0.02$, adj. $R^2 = -0.03$, $F(4,88) = 0.37$, $p = 0.83$. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.10$.

Table S1.16. First-Year Predictors of Life History Outcomes

Predictor	Outcome	N	Estimate	Std. Error	t	p (raw)	p (FDR within-predictor)	p (FDR window-wide)	Sig.
Rainfall (signed z)	Reproductive Lifespan	81	-0.91	0.62	-1.48	0.14	0.19	0.56	
Group Size	Reproductive Lifespan	81	0.41	0.55	0.75	0.45	0.91	0.94	
Maternal Demise	Reproductive Lifespan	81	-2.87	4.68	-0.61	0.54	0.69	0.94	
Mortality Rate	Reproductive Lifespan	81	0.43	1.79	0.24	0.81	0.81	0.94	
Alpha Takeovers	Reproductive Lifespan	81	-0.05	0.36	-0.12	0.90	0.90	0.94	
Fission Events	Reproductive Lifespan	81	0.66	2.17	0.30	0.76	0.86	0.94	
Rainfall (signed z)	Average Interbirth Interval	50	-48.40	31.07	-1.56	0.13	0.19	0.56	
Group Size	Average Interbirth Interval	50	-0.65	23.36	-0.03	0.98	0.98	0.98	
Maternal Demise	Average Interbirth Interval	*50*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	
Mortality Rate	Average Interbirth Interval	50	-59.79	73.75	-0.81	0.42	0.81	0.94	
Alpha Takeovers	Average Interbirth Interval	50	-17.05	15.96	-1.07	0.29	0.58	0.84	
Fission Events	Average Interbirth Interval	50	44.55	88.80	0.50	0.62	0.86	0.94	
Rainfall (signed z)	Offspring Survival to Age 4	68	-0.09	0.05	-1.78	0.08	0.19	0.56	.
Group Size	Offspring Survival to Age 4	68	-0.12	0.04	-2.73	**< 0.01**	**0.03**	0.19	**
Maternal Demise	Offspring Survival to Age 4	68	-0.51	0.35	-1.47	0.15	0.44	0.56	
Mortality Rate	Offspring Survival to Age 4	68	-0.13	0.15	-0.91	0.36	0.81	0.93	
Alpha Takeovers	Offspring Survival to Age 4	68	0.04	0.03	1.28	0.21	0.58	0.67	
Fission Events	Offspring Survival to Age 4	68	-0.30	0.16	-1.85	0.07	0.28	0.56	.
Rainfall (signed z)	Age-Expected Fertility	96	-0.06	0.15	-0.39	0.70	0.70	0.94	
Group Size	Age-Expected Fertility	96	0.03	0.14	0.21	0.83	0.98	0.94	

Predictor	Outcome	N	Estimate	Std. Error	t	p (raw)	p (FDR within-predictor)	p (FDR window-wide)	Sig.
Maternal Demise	Age-Expected Fertility	96	-0.37	0.93	-0.40	0.69	0.69	0.94	
Mortality Rate	Age-Expected Fertility	96	0.16	0.43	0.37	0.72	0.81	0.94	
Alpha Takeovers	Age-Expected Fertility	96	0.05	0.09	0.53	0.60	0.79	0.94	
Fission Events	Age-Expected Fertility	96	0.09	0.51	0.18	0.86	0.86	0.94	

Note. Linear regression results testing associations between first-year predictors and adult life history outcomes. Rainfall is coded as the signed deviation z-score (ERA5), in which positive values indicate wetter-than-average and negative values drier-than-average conditions. Maternal demise was not estimable for average interbirth interval. Model fit: Reproductive Lifespan, $R^2 = 0.07$, adj. $R^2 = -0.01$, $F(6,74) = 0.87$, $p = 0.52$; Average Interbirth Interval, $R^2 = 0.14$, adj. $R^2 = 0.04$, $F(5,44) = 1.46$, $p = 0.22$; Offspring Survival to Age 4, $R^2 = 0.19$, adj. $R^2 = 0.12$, $F(6,61) = 2.46$, $p = 0.03$; Age-Expected Fertility, $R^2 = 0.01$, adj. $R^2 = -0.06$, $F(6,89) = 0.13$, $p = 0.99$. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.10$.

Table S1.17. First-Five-Year Predictors of Life History Outcomes

Predictor	Outcome	N	Estimate	Std. Error	t	p (raw)	p (FDR within-predictor)	p (FDR window-wide)	Sig.
Rainfall (signed z)	Reproductive Lifespan	81	0.15	0.92	0.16	0.87	0.87	0.95	
Group Size	Reproductive Lifespan	81	1.14	0.86	1.32	0.19	0.19	0.63	
Maternal Demise	Reproductive Lifespan	81	-0.35	1.34	-0.26	0.80	0.89	0.95	
Mortality Rate	Reproductive Lifespan	81	2.94	3.80	0.78	0.44	0.84	0.87	
Alpha Takeovers	Reproductive Lifespan	81	-0.01	0.09	-0.12	0.91	0.91	0.95	
Fission Events	Reproductive Lifespan	81	0.81	1.16	0.70	0.49	0.68	0.87	
Rainfall (signed z)	Average Interbirth Interval	50	42.90	46.39	0.92	0.36	0.72	0.79	
Group Size	Average Interbirth Interval	50	104.48	40.06	2.61	**0.01**	**0.05**	0.30	*
Maternal Demise	Average Interbirth Interval	50	-25.21	70.24	-0.36	0.72	0.89	0.95	
Mortality Rate	Average Interbirth Interval	50	33.63	170.58	0.20	0.84	0.84	0.95	
Alpha Takeovers	Average Interbirth Interval	50	-1.28	4.34	-0.29	0.77	0.91	0.95	
Fission Events	Average Interbirth Interval	50	2.84	58.46	0.05	0.96	0.96	0.96	
Rainfall (signed z)	Offspring Survival to Age 4	68	-0.15	0.07	-2.08	**0.04**	0.17	0.33	*
Group Size	Offspring Survival to Age 4	68	0.10	0.08	1.35	0.18	0.19	0.63	
Maternal Demise	Offspring Survival to Age 4	68	0.11	0.11	1.00	0.32	0.89	0.77	
Mortality Rate	Offspring Survival to Age 4	68	-0.47	0.32	-1.47	0.15	0.59	0.63	
Alpha Takeovers	Offspring Survival to Age 4	68	0.01	0.01	1.44	0.15	0.47	0.63	
Fission Events	Offspring Survival to Age 4	68	-0.13	0.11	-1.20	0.23	0.68	0.63	
Rainfall (signed z)	Age-Expected Fertility	96	-0.08	0.21	-0.39	0.70	0.87	0.95	
Group Size	Age-Expected Fertility	96	-0.37	0.16	-2.26	**0.03**	0.05	0.32	*

Predictor	Outcome	N	Estimate	Std. Error	t	p (raw)	p (FDR within-predictor)	p (FDR window-wide)	Sig.
Maternal Demise	Age-Expected Fertility	96	0.04	0.30	0.14	0.89	0.89	0.95	
Mortality Rate	Age-Expected Fertility	96	-0.24	0.82	-0.29	0.77	0.84	0.95	
Alpha Takeovers	Age-Expected Fertility	96	0.03	0.02	1.19	0.24	0.47	0.63	
Fission Events	Age-Expected Fertility	96	-0.17	0.25	-0.67	0.51	0.68	0.87	

Note. Linear regression results testing associations between first-five-year predictors and adult life history outcomes. Rainfall is coded as the signed deviation z-score (ERA5), in which positive values indicate wetter-than-average and negative values drier-than-average conditions. Model fit: Reproductive Lifespan, $R^2 = 0.06$, adj. $R^2 = -0.01$, $F(6,74) = 0.83$, $p = 0.55$; Average Interbirth Interval, $R^2 = 0.20$, adj. $R^2 = 0.09$, $F(6,43) = 1.84$, $p = 0.11$; Offspring Survival to Age 4, $R^2 = 0.16$, adj. $R^2 = 0.07$, $F(6,61) = 1.87$, $p = 0.10$; Age-Expected Fertility, $R^2 = 0.09$, adj. $R^2 = 0.03$, $F(6,89) = 1.46$, $p = 0.20$. Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.10$.

References

Babcock, M., Wong-Parodi, G., Small, M. J., & Grossmann, I. (2016). Stakeholder perceptions of water systems and hydro-climate information in Guanacaste, Costa Rica. *Earth Perspectives*, *3*(1), 3. https://doi.org/10.1186/s40322-016-0035-x

Bateson, P., Gluckman, P., & Hanson, M. (2014). The biology of developmental plasticity and the Predictive Adaptive Response hypothesis. *The Journal of Physiology*, *592*(11), 2357–2368. https://doi.org/10.1113/jphysiol.2014.271460

Blackmore, H. L., & Ozanne, S. E. (2015). Programming of cardiovascular disease across the life-course. *Journal of Molecular and Cellular Cardiology, Perspectives on Cardiovascular Aging: Top to Bottom*, *83*, 122–130. https://doi.org/10.1016/j.yjmcc.2014.12.006

Borries, C., Larney, E., Lu, A., Ossi, K., & Koenig, A. (2008). Costs of group size: Lower developmental and reproductive rates in larger groups of leaf monkeys. *Behavioral Ecology*, *19*(6), 1186–1191. https://doi.org/10.1093/beheco/arn088

Brown, E. R., Gettler, L. T., & Rosenbaum, S. (2024). Effects of social environments on male primate HPG and HPA axis developmental programming. *Developmental Psychobiology*, *66*(5), e22491. https://doi.org/10.1002/dev.22491

Camille Melón, L., & Maguire, J. (2016). GABAergic regulation of the HPA and HPG axes and the impact of stress on reproductive function. *The Journal of Steroid Biochemistry and Molecular Biology, SI:Steroids & Nervous System*, *160*, 196–203. https://doi.org/10.1016/j.jsbmb.2015.11.019

Carnegie, S. D., Fedigan, L. M., & Ziegler, T. E. (2011). Social and environmental factors affecting fecal glucocorticoids in wild, female white-faced capuchins (Cebus capucinus). *American Journal of Primatology*, *73*(9), 861–869. https://doi.org/10.1002/ajp.20954

Charnov, E. L., & Berrigan, D. (1993). Why do female primates have such long lifespans and so few babies? *Or* Life in the slow lane. *Evolutionary Anthropology: Issues, News, and Reviews*, *1*(6), 191–194. https://doi.org/10.1002/evan.1360010604

Clarke, A. S., Wittwer, D. J., Abbott, D. H., & Schneider, M. L. (1994). Long-term effects of prenatal stress on HPA axis activity in juvenile rhesus monkeys. *Developmental Psychobiology*, *27*(5), 257–269. https://doi.org/10.1002/dev.420270502

Cottrell, E. C. (2009). Prenatal stress, glucocorticoids and the programming of adult disease. *Frontiers in Behavioral Neuroscience*, *3*. https://doi.org/10.3389/neuro.08.019.2009

Detto, M., Wright, S. J., Calderón, O., & Muller-Landau, H. C. (2018). Resource acquisition and reproductive strategies of tropical forest in response to the El Niño–Southern Oscillation. *Nature Communications*, *9*(1), 913. https://doi.org/10.1038/s41467-018-03306-9

Ellis, B. J., Figueredo, A. J., Brumbach, B. H., & Schlomer, G. L. (2009). Fundamental Dimensions of Environmental Risk. *Human Nature*, *20*(2), 204–268. https://doi.org/10.1007/s12110-009-9063-7

Ellison, P. T. (2003). Energetics and reproductive effort. *American Journal of Human Biology*, *15*(3), 342–351. https://doi.org/10.1002/ajhb.10152

Emery Thompson, M. (2017). Energetics of feeding, social behavior, and life history in non-human primates. *Hormones and Behavior, Field Endocrinology of Nonhuman Primates and Humans*, *91*, 84–96. https://doi.org/10.1016/j.yhbeh.2016.08.009

English, S., & Barreaux, A. M. (2020). The evolution of sensitive periods in development: Insights from insects. *Current Opinion in Behavioral Sciences, Sensitive and Critical Periods*, *36*, 71–78. https://doi.org/10.1016/j.cobeha.2020.07.009

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*(Suppl 1).

Fedigan, L. M., Carnegie, S. D., & Jack, K. M. (2008). Predictors of reproductive success in female white-faced capuchins (Cebus capucinus). *American Journal of Physical Anthropology*, *137*(1), 82–90. https://doi.org/10.1002/ajpa.20848

Fedigan, L. M., & Jack, K. M. (2011). Two girls for every boy: The effects of group size and composition on the reproductive success of male and female white-faced capuchins. *American Journal of Physical Anthropology*, *144*(2), 317–326. https://doi.org/10.1002/ajpa.21414

Frankenhuis, W. E., Nettle, D., & Dall, S. R. X. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 20180110. https://doi.org/10.1098/rstb.2018.0110

Gluckman, P. D., & Hanson, M. A. (2004). Developmental Origins of Disease Paradigm: A Mechanistic and Evolutionary Perspective. *Pediatric Research*, *56*(3), 311–317. https://doi.org/10.1203/01.PDR.0000135998.08025.FB

Godoy, I., Korsten, P., & Perry, S. E. (2024). Mother of all bonds: Influences on spatial association across the lifespan in capuchins. *Developmental Science*, *27*(6), e13486. https://doi.org/10.1111/desc.13486

Harvey, P. H., & Clutton-Brock, T. H. (1985). Life History Variation in Primates. *Evolution*, *39*(3), 559–581. https://doi.org/10.1111/j.1558-5646.1985.tb00395.x

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly journal of the royal meteorological society*, *146*(730), 1999-2049.

Herring, C. M., Bazer, F. W., Johnson, G. A., & Wu, G. (2018). Impacts of maternal dietary protein intake on fetal survival, growth, and development. *Experimental Biology and Medicine*, *243*(6), 525–533. https://doi.org/10.1177/1535370218758275

Huang, Y. S., Fan, C. H., Lu, Y. C., Wu, Y. T., Niu, Y. C., Hsieh, Y. H., ... & Sun, S. J. (2025). Deriving exploratory temperature-based phenological indicators under data-limited conditions: integrating ERA5 and citizen science. *Ecological Indicators*, *181*, 114484.

Humphreys, K. L., & Zeanah, C. H. (2015). Deviations from the Expectable Environment in Early Childhood and Emerging Psychopathology. *Neuropsychopharmacology*, *40*(1), 154–170. https://doi.org/10.1038/npp.2014.165

Jack, K., & Fedigan, L. M. (2004). The Demographic and Reproductive Context of Male Replacements in Cebus Capucinus. *Behaviour*, *141*(6), 755–775. https://doi.org/10.1163/1568539042245178

Jasienska, G., & Ellison, P. T. (2004). Energetic factors and seasonal changes in ovarian function in women from rural Poland. *American Journal of Human Biology*, *16*(5), 563–580. https://doi.org/10.1002/ajhb.20063

Kappeler, P. M., & Pereira, M. E. (2003). *Primate Life Histories and Socioecology*. University of Chicago Press.

Key, C., & Ross, C. (1999). Sex differences in energy expenditure in non–human primates. *Proceedings of the Royal Society B: Biological Sciences*, *266*(1437), 2479–2485. https://doi.org/10.1098/rspb.1999.0949

Lea, A. J., Altmann, J., Alberts, S. C., & Tung, J. (2015). Developmental Constraints in a Wild Primate. *The American Naturalist*, *185*(6), 809–821. https://doi.org/10.1086/681016

Lea, A. J., Tung, J., Archie, E. A., & Alberts, S. C. (2017). Developmental plasticity. *Evolution, Medicine, and Public Health*, *2017*(1), 162–175. https://doi.org/10.1093/emph/eox019

Leigh, S. R. (2004). Brain growth, life history, and cognition in primate and human evolution. *American Journal of Primatology*, *62*(3), 139–164. https://doi.org/10.1002/ajp.20012

Livadas, S., & Chrousos, G. P. (2019). Molecular and Environmental Mechanisms Regulating Puberty Initiation: An Integrated Approach. *Frontiers in Endocrinology*, *10*, 828. https://doi.org/10.3389/fendo.2019.00828

Lu, A., Petrullo, L., Carrera, S., Feder, J., Schneider-Crease, I., & Snyder-Mackler, N. (2019). Developmental responses to early-life adversity: Evolutionary and mechanistic perspectives. *Evolutionary Anthropology: Issues, News, and Reviews*, *28*(5), 249–266. https://doi.org/10.1002/evan.21791

Maestripieri, D. (2005). Effects of early experience on female behavioural and reproductive development in rhesus macaques. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1569), 1243–1248. https://doi.org/10.1098/rspb.2005.3059

McLaughlin, K. A., & Sheridan, M. A. (2016). Beyond Cumulative Risk: A Dimensional Approach to Childhood Adversity. *Current Directions in Psychological Science*, *25*(4), 239–245. https://doi.org/10.1177/0963721416655883

McLaughlin, K. A., Sheridan, M. A., Humphreys, K. L., Belsky, J., & Ellis, B. J. (2021). The Value of Dimensional Models of Early Experience: Thinking Clearly About Concepts and Categories. *Perspectives on Psychological Science*, *16*(6), 1463–1472. https://doi.org/10.1177/1745691621992346

Milton, K., & Giacalone, J. (2014). Differential effects of unusual climatic stress on capuchin (*Cebus capucinus*) and howler monkey (*Alouatta palliata*) populations on Barro Colorado Island, Panama. *American Journal of Primatology*, *76*(3), 249–261. https://doi.org/10.1002/ajp.22229

Monaghan, P. (2008). Early growth conditions, phenotypic development and environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1497), 1635–1645. https://doi.org/10.1098/rstb.2007.0011

Mumby, H. S., Mar, K. U., Thitaram, C., Courtiol, A., Towiboon, P., Min-Oo, Z., Htut-Aung, Y., Brown, J. L., & Lummaa, V. (2015). Stress and body condition are associated with climate and demography in Asian elephants. *Conservation Physiology*, *3*(1), cov030. https://doi.org/10.1093/conphys/cov030

Nettle, D., & Bateson, M. (2015). Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151005. https://doi.org/10.1098/rspb.2015.1005

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131343. https://doi.org/10.1098/rspb.2013.1343

Newell, F. L., Ausprey, I. J., & Robinson, S. K. (2023). Wet and dry extremes reduce arthropod biomass independently of leaf phenology in the wet tropics. *Global Change Biology*, *29*(2), 308–323. https://doi.org/10.1111/gcb.16379

Palacios-Vargas, J. G., Castaño-Meneses, G., Gómez-Anaya, J. A., Martínez-Yrizar, A., Mejía-Recamier, B. E., & Martínez-Sánchez, J. (2007). Litter and soil arthropods diversity and density in a tropical dry forest ecosystem in Western Mexico. *Biodiversity and Conservation*, *16*(13), 3703–3717. https://doi.org/10.1007/s10531-006-9109-7

Perry, S. (2012). The Behavior of Wild White-Faced Capuchins: Demography, Life History, Social Relationships, and Communication. In *Advances in the Study of Behavior* (Vol. 44, pp. 135–181). Academic Press. https://doi.org/10.1016/B978-0-12-394288-3.00004-6

Petersen, R. M., Patterson, S. K., Widdig, A., Turcotte, C. M., Antón, S. C., Williams, S. A., Romero, A. N., Bauman Surratt, S. E., Ruiz-Lambides, A., Cayo Biobank Research Unit, Montague, M. J., Snyder-Mackler, N., Brent, L. J. N., Higham, J. P., & Lea, A. J. (2026). Early-life adversity shapes the relationship between growth and reproduction in free-ranging female rhesus macaques. *Proceedings of the National Academy of Sciences*, *123*(19), e2524909123. https://doi.org/10.1073/pnas.2524909123

Pontzer, H. (2015). Energy Expenditure in Humans and Other Primates: A New Synthesis. *Annual Review of Anthropology*, *44*(1), 169–187. https://doi.org/10.1146/annurev-anthro-102214-013925

Pryce, C., Aubert, Y., & Maier, C. (2010). The developmental impact of prenatal stress, prenatal dexamethasone and postnatal social stress on physiology, behaviour and neuroanatomy of primate offspring: Studies in rhesus macaque and common marmoset. *Psychopharmacology*, *214*, 33–53.

Roff, D. (2001). *Life History Evolution* (1st ed.). Sinauer Associates.

Rosenbaum, S., Malani, A., Lea, A. J., Tung, J., Alberts, S. C., & Archie, E. A. (2025). Testing early life effects frameworks: Developmental constraints and adaptive response hypotheses do not explain fertility outcomes in wild female baboons. *Proceedings of the Royal Society B: Biological Sciences*, *292*(2050), 20242485. https://doi.org/10.1098/rspb.2024.2485

Sheridan, M. A., & McLaughlin, K. A. (2014). Dimensions of Early Experience and Neural Development: Deprivation and Threat. *Trends in Cognitive Sciences*, *18*(11), 580–585. https://doi.org/10.1016/j.tics.2014.09.001

Snyder, R. E., & Ellner, S. P. (2024). To Prosper, Live Long: Understanding the Sources of Reproductive Skew and Extreme Reproductive Success in Structured Populations. *The American Naturalist*, *204*(2), E11–E27. https://doi.org/10.1086/730557

Stan, K., Sanchez-Azofeifa, A., Calvo-Rodriguez, S., Castro-Magnani, M., Chen, J., Ludwig, R., & Zou, L. (2020). Climate change scenarios and projected impacts for forest productivity in Guanacaste Province (Costa Rica): Lessons for tropical forest regions. *Regional Environmental Change*, *20*(1), 14. https://doi.org/10.1007/s10113-020-01602-z

Stearns, S. C. (1976). Life-history tactics: A review of the ideas. *The Quarterly Review of Biology*, *51*(1), 3–47. https://doi.org/10.1086/409052

Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259–268. https://doi.org/10.2307/2389364

Stefanidis, K., Varlas, G., Vourka, A., Papadopoulos, A., & Dimitriou, E. (2021). Delineating the relative contribution of climate related variables to chlorophyll-a and phytoplankton

biomass in lakes using the ERA5-Land climate reanalysis data. *Water research*, *196*, 117053.

Street, M. E., Ponzi, D., Renati, R., Petraroli, M., D'Alvano, T., Lattanzi, C., Ferrari, V., Rollo, D., & Stagi, S. (2023). Precocious puberty under stressful conditions: New understanding and insights from the lessons learnt from international adoptions and the COVID-19 pandemic. *Frontiers in Endocrinology*, *14*, 1149417. https://doi.org/10.3389/fendo.2023.1149417

Taylor, S. E., Way, B. M., & Seeman, T. E. (2011). Early adversity and adult health outcomes. *Development and Psychopathology*, *23*(3), 939–954. https://doi.org/10.1017/S0954579411000411

Tilbrook, A. J. (2000). Effects of stress on reproduction in non-rodent mammals: The role of glucocorticoids and sex differences. *Reviews of Reproduction*, *5*, 105–113. https://doi.org/10.1530/ror.0.0050105

Tinsley Johnson, E., Feder, J. A., Bergman, T. J., Lu, A., Snyder-Mackler, N., & Beehner, J. C. (n.d.). The Goldilocks effect: Female geladas in mid-sized groups have higher fitness. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1952), 20210820. https://doi.org/10.1098/rspb.2021.0820

Tung, J., Archie, E. A., Altmann, J., & Alberts, S. C. (2016). Cumulative early life adversity predicts longevity in wild baboons. *Nature Communications*, *7*(1), 11181. https://doi.org/10.1038/ncomms11181

Uller, T., Feiner, N., Radersma, R., Jackson, I. S. C., & Rago, A. (2020). Developmental plasticity and evolutionary explanations. *Evolution & Development*, *22*(1–2), 47–55. https://doi.org/10.1111/ede.12314

Urlacher, S. S. (2023). The energetics of childhood: Current knowledge and insights into human variation, evolution, and health. *American Journal of Biological Anthropology*, *181*(S76), 94–117. https://doi.org/10.1002/ajpa.24719

Walker, R. S., Gurven, M., Burger, O., & Hamilton, M. J. (2008). The trade-off between number and size of offspring in humans and other primates. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1636), 827–834. https://doi.org/10.1098/rspb.2007.1511

Webster, G. D., Graber, J. A., Gesselman, A. N., Crosier, B. S., & Schember, T. O. (2014). A Life History Theory of Father Absence and Menarche: A Meta-Analysis. *Evolutionary Psychology*, *12*(2), 273–294. https://doi.org/10.1177/147470491401200202

Weibel, C. J., Tung, J., Alberts, S. C., & Archie, E. A. (2020). Accelerated reproduction is not an adaptive response to early-life adversity in wild baboons. *Proceedings of the National Academy of Sciences*, *117*(40), 24909–24919. https://doi.org/10.1073/pnas.2004018117

Welberg, L. a. M., & Seckl, J. R. (2001). Prenatal Stress, Glucocorticoids and the Programming of the Brain. *Journal of Neuroendocrinology*, *13*(2), 113–128. https://doi.org/10.1111/j.1365-2826.2001.00601.x

Wells, J. C. K. (2007). Flaws in the theory of predictive adaptive responses. *Trends in Endocrinology & Metabolism*, *18*(9), 331–337. https://doi.org/10.1016/j.tem.2007.07.006

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.

Zipple, M. N., Altmann, J., Campos, F. A., Cords, M., Fedigan, L. M., Lawler, R. R., Lonsdorf, E. V., Perry, S., Pusey, A. E., Stoinski, T. S., Strier, K. B., & Alberts, S. C. (2021). Maternal death and offspring fitness in multiple wild primates. *Proceedings of the National Academy of Sciences*, *118*(1), e2015317118. https://doi.org/10.1073/pnas.2015317118

Zipple, M. N., Archie, E. A., Tung, J., Altmann, J., & Alberts, S. C. (2019). Intergenerational effects of early adversity on survival in wild baboons. *eLife*, *8*, e47433. https://doi.org/10.7554/eLife.47433

Chapter 2: Re-evaluating the Boldness-Aggression Syndrome in a Cognitively Complex Primate

Introduction

The study of animal personality and behavioral syndromes has established that consistent individual differences in behavior are not transient noise around an adaptive optimum but are biologically meaningful targets of selection in their own right (Bell, 2007; Dingemanse & Réale, 2005; Réale et al., 2007; Sih, Bell, & Johnson, 2004; Sih, Bell, Johnson, et al., 2004; Wolf & Weissing, 2012). Animal personality is formally defined as repeatable individual differences in behavior that remain consistent across time and contexts, with repeatability *(R)* - the proportion of total phenotypic variance attributable to stable among-individual differences - serving as the core metric (Bell et al., 2009; Réale et al., 2007). When multiple such traits covary predictably at the individual level, they constitute a behavioral syndrome: an integrated suite of characteristics in which an individual's behavioral type in one domain predicts its behavior in functionally distinct domains (Sih, Bell, & Johnson, 2004). Among the behavioral syndromes documented across taxa, the boldness–aggression axis has attracted the most sustained empirical and theoretical attention (Bell, 2005; Sih, Bell, Johnson, et al., 2004). Boldness - the propensity to take risks in potentially dangerous situations - and aggression - the initiation of agonistic conflict with conspecifics - are functionally distinct but frequently correlated at the individual level, a pairing interpreted as evidence of shared proximate architecture and embedded in broader frameworks connecting behavioral strategies to life-history trajectories (Réale et al., 2007, 2010).

The conditions under which this syndrome emerges, however, are not fixed species-level properties. Experimental work has demonstrated that predation acts as an agent of correlational selection, selectively removing individuals with mismatched behavioral types and thereby

tightening the boldness–aggression correlation among survivors (Bell, 2005; Bell & Sih, 2007). This ecological contingency implies that the syndrome's strength depends on the selective landscape in which a population is embedded. A critical further insight is that population-level correlations between boldness and aggression may conceal substantial individual heterogeneity in how these traits are coupled (Dingemanse & Dochtermann, 2013). Some individuals may show strong positive coupling, others near-zero or even negative coupling, and a single population-mean correlation will be a poor summary of anyone's actual behavioral architecture. In species with complex social structures and variable life-history contexts, where the costs and benefits of boldness and aggression shift substantially across rank, age, and reproductive state, such heterogeneity may itself be biologically meaningful.

The dominant approach to studying individual differences in primate behavior has been adapted from human psychometrics, in which long-term observers rate individuals on personality items that are then subjected to factor analysis (Freeman & Gosling, 2010; Gosling, 2001). At Lomas Barbudal, this approach has identified five robust personality dimensions with high inter-rater reliability and meaningful associations with directly observed behavior and life-history outcomes (Manson & Perry, 2013; Perry et al., 2017). However, composite factors derived from factor analysis can obscure the specific behavioral covariances that define syndromes. The jingle fallacy - applying a single label to phenotypically distinct behaviors - and the jangle fallacy - using different labels for behaviors that may reflect the same underlying process - pose particular risks for syndrome research (Block, 1995; Uher, 2011a, 2011b). In the Lomas Barbudal personality data, aggression loads across both Extraversion and Neuroticism rather than cleanly onto a single factor (Manson & Perry, 2013), meaning that even a strong behavioral coupling between risk-taking and aggressiveness might not be detectable in the factor structure. These

concerns motivate a bottom-up, trait-based approach that measures specific, ethologically valid behaviors in defined functional contexts before testing their covariance (Carter et al., 2013; Hertel et al., 2020).

White-faced capuchins present a particularly interesting case for evaluating the boldness–aggression syndrome. They are among the most cognitively sophisticated New World primates, exhibiting socially learned behaviors, extractive foraging techniques, and behavioral traditions (Fragaszy et al., 2004; Perry, 2011; Perry et al., 2003; Perry, 2020). This cognitive complexity creates a theoretical tension with the proactive–reactive coping style literature, which typically associates bold and aggressive phenotypes with behavioral rigidity and routine-based strategies, and associates flexibility with reactive coping styles (Coppens et al., 2010; Koolhaas et al., 1999). For a species whose ecological success depends on behavioral invention, boldness toward novel situations may facilitate rather than impede behavioral flexibility - a prediction that inverts the standard proactive–reactive framework. Furthermore, the male reproductive career in this species places changing and often contradictory demands on the boldness–aggression profile: low-ranking males face high costs for both boldness and aggression toward dominant rivals, while alpha males may benefit from aggression toward competitors but should become risk-averse as accumulated reproductive assets increase (Clark, 1994; Stamps, 2007). The coalitionary landscape further complicates matters, because unprovoked aggression toward females may undermine the female alliances on which alpha tenure depends (Fedigan et al., 2021; Kajokaite et al., 2019; Perry, 1997, 1998, 2012). These features suggest that a rigid coupling of boldness to general aggressiveness may be maladaptive in capuchins, and that the costs and benefits of specific behavioral combinations are likely to vary substantially across individuals and life-history stages.

The developmental literature provides additional reasons to expect individual heterogeneity in syndrome expression. Early-life experience can shape both the consistency of individual behavioral traits and the degree to which they are coupled across contexts (Edenbrow & Croft, 2012; Stamps & Groothuis, 2010). State-dependent feedback models predict that initial differences in condition can be amplified through positive feedback, generating persistent individual variation in behavioral type (Houston & McNamara, 1999; Luttbeg & Sih, 2010; Stamps, 2007). If individuals differ in developmental experiences, and those experiences interact with subsequent social contexts, the resulting population may contain substantial heterogeneity in syndrome expression even when the underlying developmental processes are similar. The neuroendocrine stress response system, particularly the HPA axis, represents one pathway through which early experiences could generate lasting individual differences in both behavioral consistency and the strength of trait coupling (Sih, 2011). These developmental considerations are addressed directly in Chapter 4, which examines whether HPA axis function mediates the relationship between early-life adversity and the behavioral phenotypes characterized here.

Two specific gaps motivate the present study. First, while the psychometric personality structure of Lomas Barbudal capuchins has been thoroughly characterized (Manson & Perry, 2013), the degree to which personality ratings predict fine-grained, target-specific behavioral measures, and furthermore, the degree to which these behavioral measures themselves constitute repeatable individual traits, has not been systematically examined. Second, despite extensive theoretical interest in the boldness–aggression syndrome across taxa, it has not been rigorously tested in this population using direct behavioral observation and methods capable of detecting individual heterogeneity in behavioral coupling rather than only population-mean associations.

This study pursues three primary objectives. The first is to establish the repeatability of boldness and aggression as distinct behavioral traits in wild capuchins, assessed through direct ethological observation in specific functional contexts: snake encounter approach distances and road-crossing durations as separate boldness assays, and unprovoked agonistic behavior directed toward adult males, adult females, and non-adult conspecifics as separate aggression measures. The second is to evaluate the construct validity of observer-rated personality trait ratings by testing whether aggressiveness and fearfulness ratings predict observed behavioral rates. The third is to test whether boldness and aggression are associated on average across adult male capuchins, and to test whether the strength and direction of boldness-aggression coupling varies among individuals using both correlation-based tests and random slopes models.

I predict that both snake approach distance and road-crossing duration will show significant positive repeatability, consistent with stable individual differences in risk tolerance (Bell et al., 2009; Réale et al., 2007), but I do not assume *a priori* that the two assays will be positively correlated at the individual level; this is treated as an empirical question with the potential to reveal either domain-general boldness or context-specific risk-taking (Carter et al., 2013). For aggression, I predict significant repeatability toward each social target class, treating the degree of cross-target correlation as an empirical question. I predict that aggressiveness ratings will positively predict observed unprovoked aggression rates and that fearfulness ratings will negatively predict behavioral boldness (Manson & Perry, 2013). For syndrome structure, I predict that a population-level boldness–aggression correlation in adult males may be absent or weak, given the highly rank-dependent, contextually sensitive nature of both traits in capuchin society. However, I predict that despite the absence of a population-level syndrome, substantial individual heterogeneity in syndrome expression will be detectable through random slopes

analysis (Dingemanse & Dochtermann, 2013), with individuals varying in the degree and direction of their boldness–aggression coupling. These behavioral phenotypes - and the individual differences in their coupling - provide the foundation for Chapter 4's investigation of whether HPA axis function serves as a physiological substrate connecting early-life adversity to adult behavioral variation and life-history outcomes.

Methods

Study System

Data for the present study were drawn from the long-term behavioral database maintained by the Lomas Barbudal Monkey Project (LBMP), a study of free-living white-faced capuchin monkeys (*Cebus capucinus*) inhabiting the Reserva Biológica Lomas Barbudal and the adjoining Hacienda Pelon de la Bajura, Guanacaste Province, Costa Rica. The LBMP has collected near continuous behavioral, demographic, and life-history data on individually identified animals since 1990. Animals in the LBMP are fully habituated to human observers and are not permanently tagged or marked by researchers; individual identification is based on stable physical characteristics including facial markings, fur patterning, body condition, age, and sex. Field observers follow individually identified subjects throughout the day, recording behavioral data using standardized ethograms and scan and focal follow protocols developed and refined over three decades of continuous field operation (see Perry et al., 2012 for more detail).

Multiple social groups are monitored simultaneously. Behavioral data used in this study span the full observation history of the project, with the oldest records dating to 1990 and the most recent to 2024. Because the LBMP follows individuals throughout their lives, the dataset includes records for the same individuals across multiple life-history stages, enabling longitudinal analyses of behavioral consistency that are not possible with cross-sectional designs.

Study Subjects and Data Sources

The behavioral data analyzed in this study were derived from four sources: (1) continuous focal follow records, from which aggression measures were extracted; (2) opportunistic records of snake encounter events, from which individual approach distances were obtained; (3) road-crossing events occurring along a paved public road that bisects the home ranges of several LBMP study groups, from which crossing durations were recorded; and (4) observer trait ratings of fearfulness and aggressiveness.

For aggression analyses, the primary dataset comprised all individuals for whom a minimum of 10 focal follow segments of at least 10 minutes in length were available within a defined life-history stage (infant, early juvenile female, early juvenile male, late juvenile male, adult female, subadult male, or adult male). Because aggression is a rare event (the majority of focal follow segments contain no aggressive acts), this minimum observation threshold of 10 focal follow segments was set to ensure that per-segment aggression rates could be estimated with reasonable accuracy. The full aggression dataset comprised 580 observation rows from 340 individually identified monkeys, spanning all demographic stages and both sexes. A parallel males-only dataset was constructed by restricting to adult and subadult males (N = 166 rows, 108 individuals) to provide the focal sample for syndrome analyses and to allow estimation of repeatability within the demographic group most relevant to the study's central questions.

For boldness analyses, individuals were required to have a minimum of five observations within the relevant context (five snake encounter observations for the snake approach analysis; five road-crossing events for the road-crossing analysis). This threshold was selected as the minimum necessary to obtain stable individual-level estimates (best linear unbiased predictors, BLUPs) from the mixed-effects models used to extract individual boldness scores. Because road-

crossing data are available only for individuals whose home ranges overlap the paved road, and snake encounter data depend on encounter occurrence within the observation period, the samples for these analyses are smaller and differ in demographic composition from the aggression sample. The snake boldness sample available for syndrome analyses comprised 29 adult and subadult males; the road boldness sample comprised 28 adult and subadult males. Sensitivity analyses using a relaxed minimum threshold of two observations are reported where relevant to assessing the robustness of findings to this choice.

For personality trait rating analyses, ratings from the LBMP's long-term psychometric assessment protocol (described below) were available for a subset of 353 observation rows (reflecting individuals at distinct life stages/age classes) from 280 individuals, including both sexes and all demographic stages. Individuals were required to have a minimum of two raters providing a score during the relevant demographic stage to be included in rating-based analyses; sensitivity analyses using the full available rating data without this restriction are reported separately.

Behavioral Measures

Aggressive Behavior

Rates of aggressive behavior were derived from 10-minute continuous focal animal follows (Altmann, 1974) conducted as part of the LBMP's standard data collection protocol. During each focal follow, field observers recorded all dyadic social interactions involving the focal individual, including the identity of the interaction partner and the directional nature of any agonism. The behavioral ethogram used to identify and code aggression has been standardized across the project's history and includes the following acts initiated by the focal individual: bite, chase, hit, bounce, pull, nip, snap at, push, lunge, glare, threat face, wrestle, pounce, swipe at,

and vocal threat. A focal follow segment is defined as a single, uninterrupted 10-minute observation period. The number of follow segments per individual varies as a function of their time in the dataset and the intensity of observer effort during each period.

Three primary measures of unprovoked agonistic behavior were analyzed, each representing aggression directed toward a distinct social target class: (1) unprovoked aggression toward adult males, (2) unprovoked aggression toward adult females, and (3) unprovoked aggression toward non-adult conspecifics (juveniles and infants). An aggressive bout was classified as unprovoked if it was initiated by the focal individual without any immediately preceding agonistic act by the target individual within the preceding minute directed at the focal animal. Aggression toward each target class was summarized as a count of unprovoked aggressive acts within each observation stage, paired with the total number of focal follow segments observed during that stage.

The decision to treat aggression toward different social targets as functionally distinct behavioral measures, rather than aggregating them into a single overall aggression rate, reflects the theoretical argument that the costs, benefits, and motivational substrates of agonism vary substantially depending on the identity of the target. Unprovoked aggression toward an adult male (particularly a dominant male) carries qualitatively different risks and potential payoffs than equivalent behavior toward a female or a juvenile, and these differences have well-documented consequences for rank dynamics and social relationships in this population (Gros-Louis et al., 2003; Perry et al., 2017). Treating these as separate behavioral modules allows the analysis to test whether individual consistency in aggression is general across targets or whether it is expressed selectively, a distinction that is informative about the structure of the behavioral phenotype.

Snake Encounter Boldness

Boldness in a predator-relevant context was quantified using observational data from snake encounters occurring naturally within the home ranges of LBMP study groups. Two snake species known to pose genuine predation risk to capuchins were included in all primary analyses: the Central American rattlesnake (*Crotalus simus*) and the boa constrictor (*Boa constrictor*). When a snake was detected by the group, the snake became the focal subject of an event-based observation record, during which the approach distances of all attending individuals were recorded in body lengths - the distance between the attending individual and the snake, expressed as a multiple of the monkey's own estimated body length. Because multiple individuals typically attend a snake encounter, simultaneous observations (individual approach distances) are nested within snake encounter events, which are in turn nested within individual monkeys (each monkey may attend multiple encounters across its observational history). Observations from snakes not belonging to the two focal dangerous species were excluded from all analyses on the grounds that they do not provide a biologically valid measure of individual risk tolerance.

Snake approach data is drawn from two largely distinct periods of data collection within the project history of the LBMP. During the period spanning 2002-2005, distances were collected in binned intervals of 1, 5, or 10 body lengths. This collection protocol was used across 3503 observations. From 2006-2024, a continuous numeric scale was used instead, allowing approach distances to vary from 0-20+ body lengths, and resulting in an additional 1602 observations. To account for variation in precision due to the use of different data collection protocols, a separate variable – precision – indicating whether the data collected was categorically binned or continuous was added to all models.

Approach distance was log-transformed prior to analysis to improve distributional properties and to reflect the non-linear scaling of perceived ecological risk with physical proximity. Approaching from ten to nine body lengths carries qualitatively different threat implications than approaching from two to one body length. The primary model used for repeatability estimation and BLUP extraction included log-transformed approach distance as the response variable with age class, precision, and snake species (rattlesnake vs. boa) as fixed effects, and both individual monkey identity and snake encounter event identity as crossed random intercepts. Including snake event as a random intercept partitions variance attributable to encounter-specific factors - such as snake position, activity state, size, or group composition on a given day - from the variance attributable to consistent individual tendencies, which is the quantity of interest for both repeatability and BLUP estimation. Individual boldness scores were extracted as BLUPs for the individual-identity random effect. Because the BLUP represents deviation from the population mean in the direction of greater approach distance (i.e., less bold behavior), scores were sign-reversed prior to downstream analyses so that higher values consistently indicate greater boldness.

Road-Crossing Boldness

Boldness in an anthropogenic risk context was quantified using the duration, in seconds, of road-crossing events occurring on a paved public road that bisects the home ranges of several LBMP study groups. The road represents an ecologically valid risk-taking challenge: vehicle traffic on this road has been documented as a source of mortality in this population, and crossing events are readily observable and can be timed precisely by field observers. Individuals vary substantially in their crossing behavior, from rapid sprints lasting a couple of seconds to prolonged, interrupted crossings involving multiple pauses on the road surface.

Road-crossing durations were recorded for all observed crossing events for all study subjects whose home ranges overlapped with the relevant road segment. Observations with crossing durations exceeding 600 seconds were excluded prior to analysis on the grounds that they most likely reflect errors in data transcription rather than genuine individual differences in risk tolerance. This applied to 10 out of 5971 observations.

Crossing duration was log-transformed to address the pronounced right skew in raw crossing times. The primary model for repeatability estimation and BLUP extraction included log-transformed crossing duration as the response variable with age class and car presence (a binary indicator of whether a vehicle passed during the crossing event) as fixed effects and individual monkey identity as a random intercept. Car presence was included because vehicles plausibly influence crossing speed irrespective of an individual's baseline boldness tendencies and including it as a covariate improves the precision of individual-level estimates. Individual boldness BLUPs were extracted from this model. Because longer crossing durations indicate greater time spent in a risky location, higher BLUP values are interpreted as indicating greater boldness, and no sign reversal was applied.

Personality Trait Ratings

Personality trait ratings were available for a subset of LBMP individuals through the project's long-term psychometric assessment protocol. Beginning in 2002, observers with at least one year of continuous fieldwork experience with the LBMP - including field assistants, graduate students, postdoctoral researchers, and principal investigators - have been asked to complete a 26-item capuchin-specific personality questionnaire for individuals they know well through direct field observation. Occasional exceptions to this rule werere made for observers who spent less time on site but focused on very few groups (or even one group), and who therefore had

adequate knowledge of the groups they were rating. The questionnaire was developed and validated specifically for this population and is described in detail in (Manson & Perry, 2013). Inter-rater reliability is high across the majority of items, and 15 personality facets have been shown to be stable across adolescence and adulthood in this population (Manson & Perry, 2013).

Two personality facets were of primary interest in the present study. The aggressiveness facet (inter-rater reliability: ICC(3,k) = 0.99, calculated within this study sample) was used as a predictor of observed aggression rates in the personality rating validity analyses. The fearfulness facet (ICC(3,k) = 0.97 , calculated within this study sample) was used as a proxy for boldness. We predicted a negative association between fearfulness ratings and behavioral boldness, reasoning that individuals perceived by observers as more fearful would be less likely to engage in risky behavior.

Mean trait ratings were calculated for each individual within each demographic stage by averaging across all raters who provided a score during that period. This stage-stratified averaging approach was adopted to account for the possibility that observers' perceptions of an individual's personality may track genuine behavioral changes across life-history stages, a known feature of this population (Manson & Perry, 2013), rather than treating ratings collected during different life stages as equivalent measures of a single, static trait and also to allow for proper alignment with the road crossing and snake approach behavioral assay such that personality ratings of adult individuals are not being used to predict infant behavior or vice versa. The aggressiveness rating was used as a continuous predictor of observed aggression, and the fearfulness rating was used as a continuous predictor of boldness. Ratings were standardized (z-scored) within each analysis to facilitate comparison of effect sizes across models.

Statistical Analyses

All analyses were conducted in R version 4.3. Mixed-effects models were fit using the lme4 package (Bates et al., 2015), with significance of random effects assessed via likelihood ratio tests using the ranova() function in lmerTest (Kuznetsova et al., 2017).

Objective 1: Individual Variance in Aggression and Target-Specificity

To assess whether individual capuchins differ in their propensity to aggress across social targets, and whether such differences are general or target-specific, I estimated the proportion of variance in unprovoked aggression toward each of the three target classes (adult males, adult females, non-adults) using binomial generalized linear mixed-effects models (GLMMs). Because aggression is a rare event - the majority of focal follow segments contain no aggressive acts - the response in each model was specified as a binomial outcome: the count of follow segments during which the focal individual initiated unprovoked aggression toward the target class (number of "successes") out of the total number of follow segments observed in the relevant life-history stage (number of "trials"). This formulation appropriately accounts for variation in observation effort across individuals and stages and avoids the assumption violations that arise from modeling bounded proportions or zero-inflated count data with a Gaussian error structure.

All models included age class as a fixed effect (with infants as the reference category) and alpha status (coded as a binary covariate: 1 = settled alpha tenure, 0 = non-alpha or uncertain) as an additional fixed covariate. Periods of contested or uncertain dominance were coded as non-alpha. Individual monkey identity was included as a random intercept in all models to partition stable among-individual variance in the propensity to aggress from within-individual (residual) variance.

Two parallel model sets were fit: a full model including all age classes and both sexes, and a males-only model restricted to adult and subadult males. The males-only models were the

primary focus for syndrome analyses (see below) and provided repeatability estimates within the most demographically homogeneous and theoretically relevant subgroup. The full model was included to characterize the developmental trajectory of aggression across the lifespan and to verify that age class was functioning as expected as a covariate.

Repeatability (R) was estimated from each model as the intraclass correlation coefficient (ICC) on the latent logit scale. The ICC was calculated as:

$$R = \sigma^2\_ind / (\sigma^2\_ind + \sigma^2\_\varepsilon)$$

where σ^2_ind is the among-individual variance estimated from the random intercept and σ^2_ε is the residual variance on the latent scale, approximated as $\pi^2/3 \sim 3.29$ for binomial models with a logit link (Nakagawa & Schielzeth, 2010). Bootstrap confidence intervals (95%) around each repeatability estimate were obtained via parametric bootstrap using the bootMER package (Bates et al., 2015) with 1,000 bootstrap replicates using the same model formula as the primary analysis.

To characterize the structure of aggression across social targets – specifically, whether individual differences in aggressiveness are general across targets or target-specific -– Spearman rank correlations were computed between all pairwise combinations of per-segment aggression rate measures within the adult male subsample.

Objective 2: Personality Rating Validity

To evaluate the construct validity of the LBMP's psychometric personality ratings, we tested whether human-assigned ratings of aggressiveness predicted observed aggression rates across social target classes, and whether fearfulness ratings predicted individual boldness estimates. These analyses ask whether the personality dimensions captured through observer

judgment correspond to variation in the specific, ethologically defined behavioral measures used in this study.

Prior to computing mean trait ratings, rater reliability was assessed using two complementary approaches. First, inter-rater reliability for the two focal trait dimensions – aggressiveness and fearfulness – was estimated using the intraclass correlation coefficient ICC(3,k). Second, individual rater consistency was evaluated using a leave-one-out (LOO) peer disagreement procedure to identify raters whose ratings were systematically erratic relative to their peers (for more, see Supplementary Materials). For aggressiveness ratings, a series of binomial GLMMs was fit following the same structure as the repeatability models described above – binomial response, age class as a fixed covariate, individual monkey identity as a random intercept – with mean aggressiveness rating (standardized to zero mean and unit standard deviation) as the key predictor of interest. Age class was retained as a fixed covariate in these models not only to account for developmental variation in aggression rates, but also because observer ratings may reflect preformed expectations about aggressiveness that co-vary with an individual's perceived age-sex class independently of their true behavioral tendencies. Including age class as a covariate therefore provides a more stringent test of whether personality ratings capture genuine individual-level behavioral variation beyond what could be inferred from demographic category membership alone. Separate models were fit for each of the three social target classes. These models were run on the full dataset restricted to individuals with available personality ratings (N = 353 rows, 280 individuals).

For fearfulness ratings, Spearman's rank correlations were computed between mean fearfulness ratings and individual boldness BLUPs from the snake approach and road-crossing models, across the full sample of individuals for whom both behavioral boldness estimates and

personality ratings were available, without restriction by sex or age class. Sensitivity analyses using a relaxed minimum observation threshold ($n \geq 2$ behavioral observations and no minimum rater requirement) were conducted to assess robustness.

Boldness Repeatability and Cross-Context Consistency

Repeatability of snake approach boldness and road-crossing boldness was estimated separately from the primary mixed-effects models described under Behavioral Measures above. For snake approach behavior, repeatability was estimated as the proportion of total variance in log approach distance attributable to individual monkey identity:

$$R = \sigma^2_{ind} / (\sigma^2_{ind} + \sigma^2_{event} + \sigma^2_{residual})$$

where σ^2_{ind} is the variance of the individual-identity random intercept, σ^2_{event} is the variance of the snake encounter event random intercept, and $\sigma^2_{residual}$ is the within-individual residual variance. For road-crossing behavior, variance was partitioned between the individual random intercept and the within-individual residual (no event-level nesting was present in the road-crossing design):

$$R = \sigma^2_{ind} / (\sigma^2_{ind} + \sigma^2_{residual})$$

Analyses were conducted on log-transformed response variables, which provided substantially better fit than untransformed models in both behavioral contexts. Bootstrap confidence intervals (95%, 1,000 replicates) were obtained using the bootMER package (Bates et al., 2015). The relationship between individual boldness (BLUP estimate) and within-individual behavioral flexibility was also examined. Flexibility was operationalized as the within-individual standard deviation of residuals from each individual's predicted behavior based on fixed effects alone (i.e., the deviation of each observation from the fixed-effects prediction, excluding the

individual's random intercept). For each boldness context, these residuals were computed within individual and demographic stage, and flexibility was the standard deviation of those residuals. The Spearman rank-order correlation between boldness BLUPs and flexibility scores was then computed within each context to test whether bolder individuals are also more variable in their behavior - a relationship that would be inconsistent with the traditional characterization of bold, proactive phenotypes as behaviorally rigid.

To test whether boldness in the snake context and boldness in the road-crossing context reflect the same underlying individual trait, Spearman's rank correlation was computed between the two sets of BLUP estimates across the full sample of individuals with data in both contexts, and separately within the subset of adult and subadult males. Because the two boldness measures are derived from qualitatively different ecological contexts - predator-relevant approach to a live animal versus anthropogenic risk during a road crossing - a positive cross-context correlation would constitute evidence for a domain-general boldness construct, while a null result would be consistent with context-specific risk-taking modules that do not reflect a unified latent trait.

Objective 3: Boldness–Aggression Behavioral Syndrome

Two complementary analytical approaches were used. The primary analysis was a random slopes mixed-effects model that simultaneously estimated the average boldness–aggression association at the population level and the variance in that association across individuals. Data were restructured into long format such that each row represented a unique combination of individual, boldness context (snake or road), and aggression context (unprovoked toward adult males, unprovoked toward adult females, or unprovoked toward non-adults), yielding a data structure in which each individual contributes multiple rows corresponding to

different combinations of their boldness estimate and their aggression rates across different target classes. The model was specified as:

aggression ~ boldness + aggression_context + (1 + boldness | monkey_id)

where boldness (the individual's context-specific BLUP score) and aggression_context (a factor distinguishing the three target classes) were fixed effects, and both a random intercept and a random slope for boldness were estimated for each individual monkey. The fixed effect of boldness estimates the *average* association between boldness and aggression across individuals and aggression contexts - the population-level syndrome test. The random slope for boldness captures individual variation in the boldness–aggression relationship: a significant random slope variance, assessed via a likelihood ratio test comparing the model with and without the random slope term (ranova() function in lmerTest), indicates that individuals differ significantly in how strongly their boldness predicts their aggression, regardless of whether the population-mean slope differs from zero.

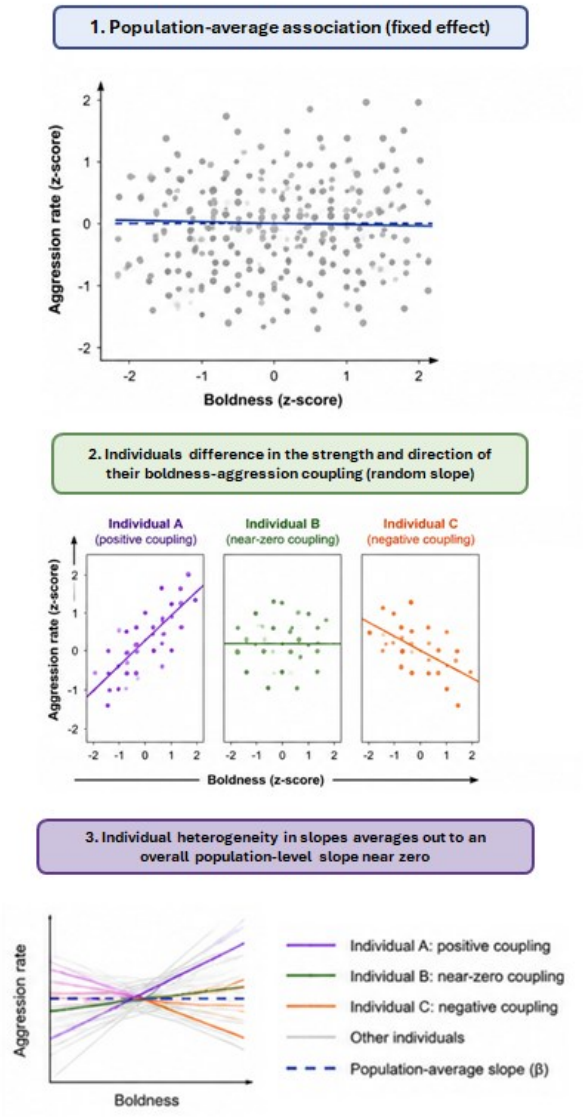


Figure 2.2 Conceptual diagram of possible relationship between populatio-average association between boldness and aggression and individual heterogeneity in behavioral coupling.

The distinction between these two tests is important. A null fixed effect of boldness indicates that there is no *average* boldness–aggression coupling across the population, that is, knowing an individual is bold does not, on average, improve prediction of their aggression level. A significant random slope variance, by contrast, indicates that there is substantial *individual heterogeneity* in syndrome expression - that some individuals show strong positive coupling, others near-zero coupling, and possibly others negative coupling - and that this variation is not reducible to chance. These two findings can co-occur, and their co-occurrence would suggest that a behavioral syndrome exists as an individually variable phenomenon rather than as a uniform population-level characteristic (see Figure 2.1).

The random slopes model was fit separately for adult males (the primary focal group), adult females, and juveniles (the latter two as exploratory comparisons). Models returning singular fits for the female and juvenile subgroups are noted and interpreted conservatively. For each demographic group, the distribution of individual-specific slope estimates was extracted and visualized by plotting each individual's

estimated boldness–aggression slope in rank order. This visualization directly displays the extent and pattern of heterogeneity in syndrome expression across individuals.

As a supplementary analysis to facilitate comparison with prior literature that has tested behavioral syndrome structure using simple bivariate correlations, Spearman's rank correlations were computed between each boldness BLUP (snake and road, separately) and each logit-transformed aggression rate (toward adult males, adult females, and non-adults) within the sample of males with both boldness and aggression data, yielding six correlation coefficients. Aggression data were aggregated to the individual level by summing counts of aggressive acts and total follow segments across all observation periods within the adult and subadult male stages, and per-segment aggression probabilities were logit-transformed (with a continuity correction of 0.001 added to zero values prior to transformation) to improve distributional properties. All six p-values were adjusted using the false discovery rate (FDR) correction. A post-hoc power analysis was conducted using the pwr package (Champely, 2020) to determine the minimum detectable effect size (Spearman's ρ) at 80% power and $\alpha = 0.05$ given the observed sample sizes.

As an exploratory supplement, latent profile analysis (LPA) was conducted on the expanded sample of individuals with both snake boldness BLUPs and aggression data ($N = 112$, including both sexes and all age classes with available data in both measures), using the tidyLPA package (Rosenberg et al., 2018). Road crossing boldness was not included as a second boldness variable both because it demonstrates considerably lower repeatability than the snake assay, and also because the subset of individuals with complete data in both boldness contexts is too small to support stable multi-variable profile solutions. LPA models with two, three, and four latent profiles were estimated under an equal-variance, zero-covariance parameterization. Solutions

were compared using the Bayesian Information Criterion (BIC) and mean posterior classification probabilities. LPA was included as an exploratory tool to assess whether the continuous distribution of individual syndrome expression is better characterized by a small number of discrete behavioral types than by a continuum. A finding of poor classification accuracy or BIC preference for a single-profile solution would support the continuous heterogeneity interpretation favored by the random slopes analysis.

Results

Repeatability of Aggression and Boldness as Distinct Behavioral Traits

Individual Differences in Aggression and Target-Specificity

Across all six binomial GLMMs - three target classes fitted across the full and males-only datasets - alpha male status did not significantly predict aggression rates toward any social target class in either dataset (all $p > 0.06$). The one near-significant trend was a marginal tendency for alpha males to show higher female-directed aggression (full model: $\beta = 0.29$, SE = 0.16, $z = 1.82$, $p = 0.p = 0.069$; males-only: $\beta = 0.29$, SE = 0.16, $z = 1.77$, $p = 0.p = 0.076$), which did not survive conventional significance thresholds. Aggression showed a consistent and pronounced developmental trajectory across all three target classes: rates increased significantly from infancy through adulthood in both models (all age class coefficients $p < 0.05$, excepting early juvenile males in the full model), suggesting that developmental stage is a primary driver of variation in observed aggression rates.

ICC estimates were low across all target classes and both model sets (Table 2.1). In the full model, ICC values ranged from $R = 0.036$ for non-adult-directed aggression to $R = 0.074$ for female-directed aggression. ICC estimates were uniformly higher in the males-only models,

ranging from $R = 0.048$ to $R = 0.087$, consistent with the expectation that restricting to a more demographically homogeneous subsample reduces between-class variance that would otherwise dilute individual signal. All estimates fall substantially below the meta-analytic mean of $R \sim 0.37$ reported by (Bell et al., 2009), suggesting that aggression in white-faced capuchins is more situationally labile than the cross-taxa average.

Individual identity accounted for the greatest proportion of variance in female-directed aggression in both the full (ICC = 0.074) and males-only (ICC = 0.087) models. Individual identity accounted for the least variance in non-adult-directed aggression in the full model (ICC = 0.036), consistent with this target class representing the least competitively consequential context and therefore the most situationally driven form of agonism. Spearman rank correlations between per-segment aggression rates toward the three social target classes revealed no evidence of a generalized aggressiveness dimension (Table 2.2). Male-directed and female-directed aggression rates were essentially uncorrelated ($\rho = -0.014$, 95% CI [−0.166, 0.139], p = 0.p = 0.858), as were male-directed and non-adult-directed rates ($\rho = 0.034$, 95% CI [−0.119, 0.185], p = 0.p = 0.668). A modest positive association was observed between female-directed and non-adult-directed aggression rates ($\rho = 0.244$, 95% CI [0.096, 0.382], p = 0.p = 0.002), suggesting some overlap in the tendency to aggress toward subordinate or low-threat targets. Critically, the absence of any positive correlation between male-directed aggression and either other target class (and the near-zero relationship between the two most functionally distinct targets) is inconsistent with a single aggressiveness trait and consistent with target-specific behavioral regulation in which competitive agonism toward rivals and social agonism toward females and non-adults represent partially independent behavioral dimensions.

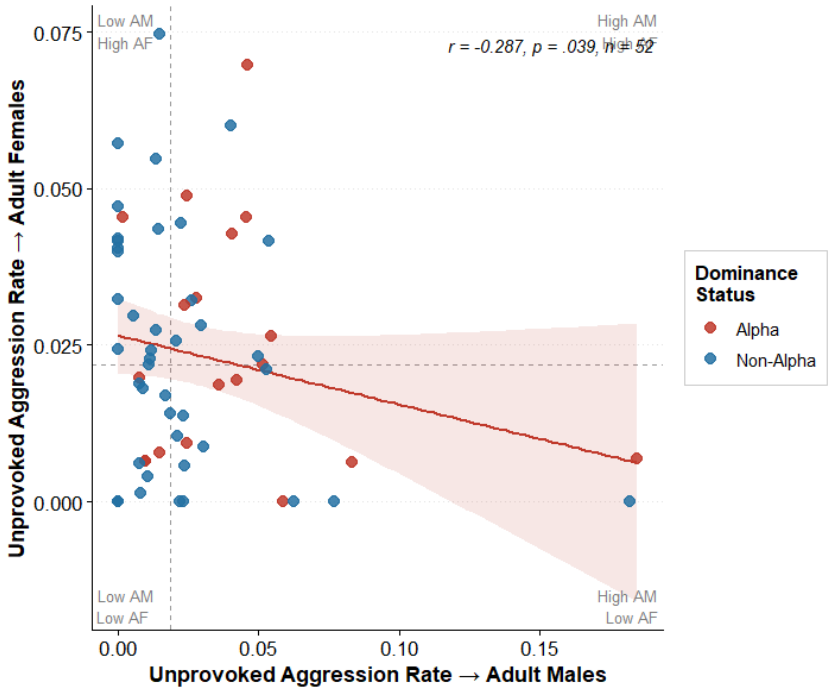


Figure 2.2. Target-specificity of aggression in adult male capuchins. Each point represents one adult male (N = 52). A significant negative correlation indicates that males who are more aggressive toward other males tend to be less aggressive toward females, consistent with target-specific rather than general aggressiveness. Dashed lines indicate median splits. Points colored by alpha status.

Table 2.1. ICC estimates with 95% bootstrap confidence intervals (1,000 replicates) for unprovoked aggression toward each social target class, derived from full (all age classes and sexes; N = 580 rows, 340 individuals) and males-only (adult and subadult males; N = 166 rows, 108 individuals) binomial GLMMs. All models included age class and alpha status as fixed effects and individual monkey identity as a random intercept.

Target class	**Model**	**ICC**	**95% CI**	**p (LRT)**
Unprovoked → adult males	Full	0.048	[0.031, 0.064]	**<0 .001**
	Males only	0.076	[0.04 – 0.11]	< 0.001
Unprovoked → adult females	Full	0.074	[0.074 – 0.049]	< 0.001
	Males only	0.087	[0.047 – 0.126]	< 0.001

Unprovoked → non-adults	Full	0.036	[0.025 – 0.047]	< 0.001
	Males only	0.048	[0.025 – 0.07]	< 0.001

Note. CI = confidence interval estimated via parametric bootstrap (bootMER; Bates, 2017). LRT = likelihood ratio test for the random intercept variance component. ICC estimated on the latent logit scale (Nakagawa & Schielzeth, 2010).

Table 2.2. Spearman's bivariate rank correlations between rates of male-directed unprovoked aggression towards distinct social targets (n = 63).

Target classes	**ρ**	**95% CI**	**P**
Adult males & Adult females	-0.114	[-0.365, 0.152]	0.372
Adult males & non-adults	0.116	[-0.151, 0.366]	0.367
Adult females & non-adults	0.333	[0.078, 0.547]	0.008

Personality Rating Validity: Aggressiveness

Rated aggressiveness significantly predicted observed unprovoked aggression toward all three social target classes in the full binomial GLMM (N = 370 rows, 285 individuals; Table 3). A one standard deviation increase in aggressiveness rating was associated with a significant increase in the log-odds of a focal follow segment containing aggression toward adult males (β = 0.179, SE = 0.047, $z = 3.78$, $p < 0.001$, 95% CI [0.087, 0.271]), adult females (β = 0.325, SE = 0.063, $z = 5.16$, $p < 0.001$, 95% CI [0.202, 0.449]), and non-adult conspecifics (β = 0.187, SE = 0.042, $z = 4.43$, $p < 0.001$, 95% CI [0.105, 0.270]).

The predictive validity of aggressiveness ratings was strongest for female-directed aggression (β = 0.325) and approximately equal for male-directed and non-adult-directed aggression (β ~ 0.19). Significant positive prediction across all three target classes establishes convergent validity between the psychometric rating instrument and directly observed behavioral

measures, indicating that observer-assigned aggressiveness scores capture genuine and generalizable individual differences in agonistic behavior. The relatively stronger association with female-directed aggression may reflect that conspicuous female-directed agonism is among the most perceptually salient behavioral signals available to long-term observers, making it the primary behavioral basis for their aggressiveness judgments.

Table 2.3. Effects of standardized aggressiveness trait rating on observed unprovoked aggression toward each social target class, estimated from binomial GLMMs including age class as a fixed covariate and individual monkey identity as a random intercept (N = 370 rows, 285 individuals). All models converged without singular fits.

Outcome	**β**	**SE**	**z**	**p**	**95% CI**
Unprovoked → adult males	0.179	0.047	3.78	< 0.001	[0.087, 0.271]
Unprovoked → adult females	0.325	0.063	5.16	< 0.001	[0.202, 0.449]
Unprovoked → non-adults	0.187	0.042	4.42	< 0.001	[0.105, 0.270]

Note. β = log-odds coefficient for a one standard deviation increase in aggressiveness rating. SE = standard error. 95% CI in brackets. All models fit using bobyqa optimizer.

Boldness Repeatability: Snake Encounter Context

Mixed-effects model comparisons on log-transformed approach distances from dangerous snake encounters (N = 4,206 observations; 158 individuals; 513 encounters) showed that the model including both age class and snake species as fixed effects (AIC = 5,529.5) provided substantially better fit than models including only age class (ΔAIC = 26.0) or the intercept-only baseline (ΔAIC = 79.0). This best-fitting model was used for all subsequent variance partitioning and BLUP extraction.

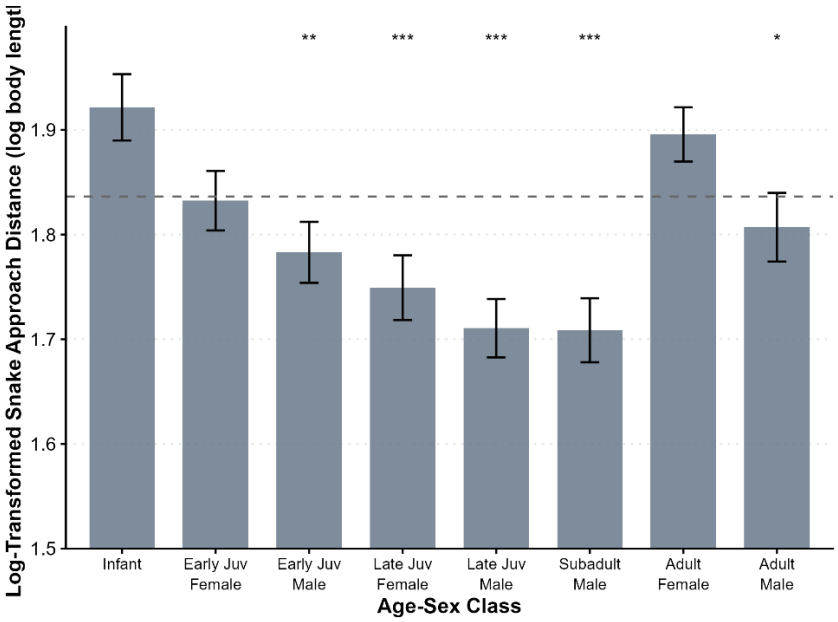


Figure 2.3. Developmental variation in snake approach behavior**.** All age-sex classes approach significantly closer than adult females (the reference group) except infants (β = −0.018, p = 0.p = 0.929), with subadult males (β = −1.132, p < 0.001) and late juvenile males (β = −1.138, p < 0.001) approaching most closely. Adult males also approach significantly closer than adult females (β = −0.584, p = 0.p = 0.007). Asterisks indicate significant difference from adult females: *p < 0.05, **p < 0.01, ***p < 0.001.

Age class was a highly significant predictor of approach distance. Relative to the adult female reference group, subadult males approached most closely (β = −0.185, SE = 0.033, $p < 0.001$), followed by late juvenile males (β = −0.181, SE = 0.031, $p < 0.001$), early juvenile males (β = −0.102, SE = 0.032, $p = 0.p = 0.001$), and adult males (β = −0.090, SE = 0.035, $p = 0.p = 0.011$). Infants did not differ significantly from adult females (β = 0.040, SE = 0.033, $p = 0.p = 0.236$). Cascabel rattlesnakes elicited significantly greater approach caution than boa constrictors (β = 0.258, SE = 0.038, $p < 0.001$), indicating that species identity meaningfully shaped approach behavior.

Variance partitioning revealed a striking dominance of encounter-specific over individual-level variance. Of total variance in log approach distance, 44.2% was attributable to differences between snake encounter events ($\sigma^2_{event} = 0.156$), 54.1% to within-individual residual variance ($\sigma^2_{resid} = 0.164$), and only 1.7% to stable individual differences ($\sigma^2_{ind} = 0.007$), yielding

a repeatability estimate of $R = 0.023$ (95% CI [0.022, 0.059]). This value is roughly 18-fold lower than the cross-taxa meta-analytic mean (Bell et al., 2009), indicating that individual differences in snake-context boldness are not meaningfully stable across encounters. The dominant role of encounter-level variance suggests that the particular characteristics of each snake event explain far more of the variation in approach behavior than consistent individual tendencies do.

Despite this low repeatability, bolder individuals (those whose BLUPs indicated closer approach than expected given age class and snake species) showed significantly greater within-individual variability in their approach distances (Spearman's $\rho = 0.264$, 95% CI [0.138, 0.381], $p < 0.001$, $n = 226$). This positive boldness–flexibility relationship - whereby behaviorally bolder individuals also show more variable responses across encounters - is the opposite of the rigid-proactive phenotype predicted by standard proactive–reactive coping style frameworks, and suggests that in this context, boldness may enable behavioral exploration rather than constrain it.

Boldness Repeatability: Road-Crossing Context

Sequential model comparisons on log-transformed road-crossing durations (N = 5,960 observations; 118 individuals) identified the model including both age-sex class and car presence as fixed effects as the best fit ($\chi^2(7) = 102.04$, $p < 0.001$ improvement over the car-presence-only model; interaction term did not improve fit, $\chi^2(7) = 7.92$, $p = 0.340$). Car presence was associated with a 5% increase in crossing duration on the log scale ($\beta = 0.051$, SE = 0.023, $p = 0.025$). Age class effects were pronounced: adult males crossed approximately 51% more slowly than adult females ($\exp(0.413) = 1.51$, $p < 0.001$), subadult males crossed 13% more slowly ($\exp(0.125) = 1.13$, $p = 0.067$, marginal), and early juvenile females crossed 30% faster ($\exp(-0.356) = 0.70$, $p < 0.001$).

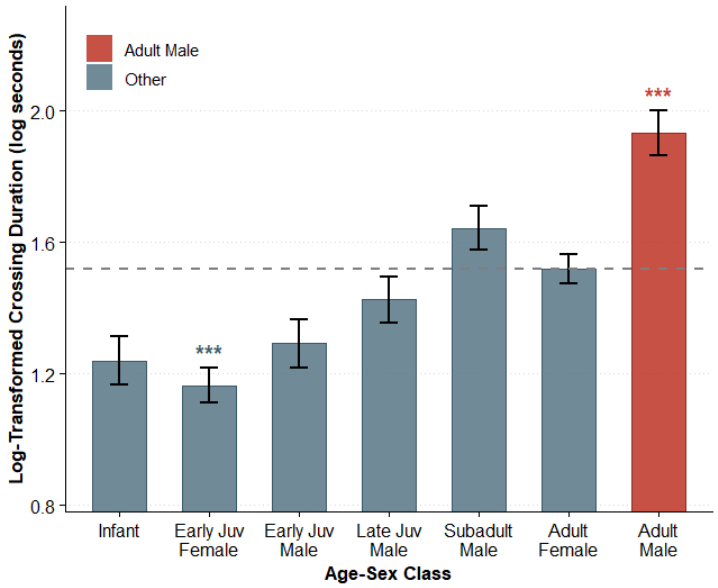


Figure 2.4: Developmental variation in road crossing behavior. Adult males cross significantly more slowly than other age classes ($\beta = 0.413$, $p < 0.001$), while early juvenile females cross fastest ($\beta = -0.356$, $p < 0.001$)

Variance partitioning from the log-transformed model yielded a repeatability of R = 0.070 ($\sigma^2_{ind} = 0.042$, $\sigma^2_{resid} = 0.552$; 95% CI [0.0433, 0.100]), approximately four times higher than that observed in the snake context and representing modest but non-trivial individual consistency. As with the snake context, this value remains substantially below the cross-taxa meta-analytic benchmark, suggesting that road-crossing behavior is also highly situationally variable, though to a lesser degree than snake-approach behavior.

The positive boldness–flexibility relationship observed in the snake context replicated in the road-crossing context. Bolder individuals (those with higher road-crossing BLUPs, indicating longer crossing durations) showed significantly greater within-individual variability in crossing duration (Spearman's $\rho = 0.288$, 95% CI [0.153, 0.413], $p = < .001$, $n = 191$). The convergence of this relationship across two functionally distinct risk contexts - predator encounter and vehicle traffic - suggests that the positive association between boldness and behavioral flexibility is a consistent feature of the behavioral phenotype in this population rather than a context-specific

artifact.

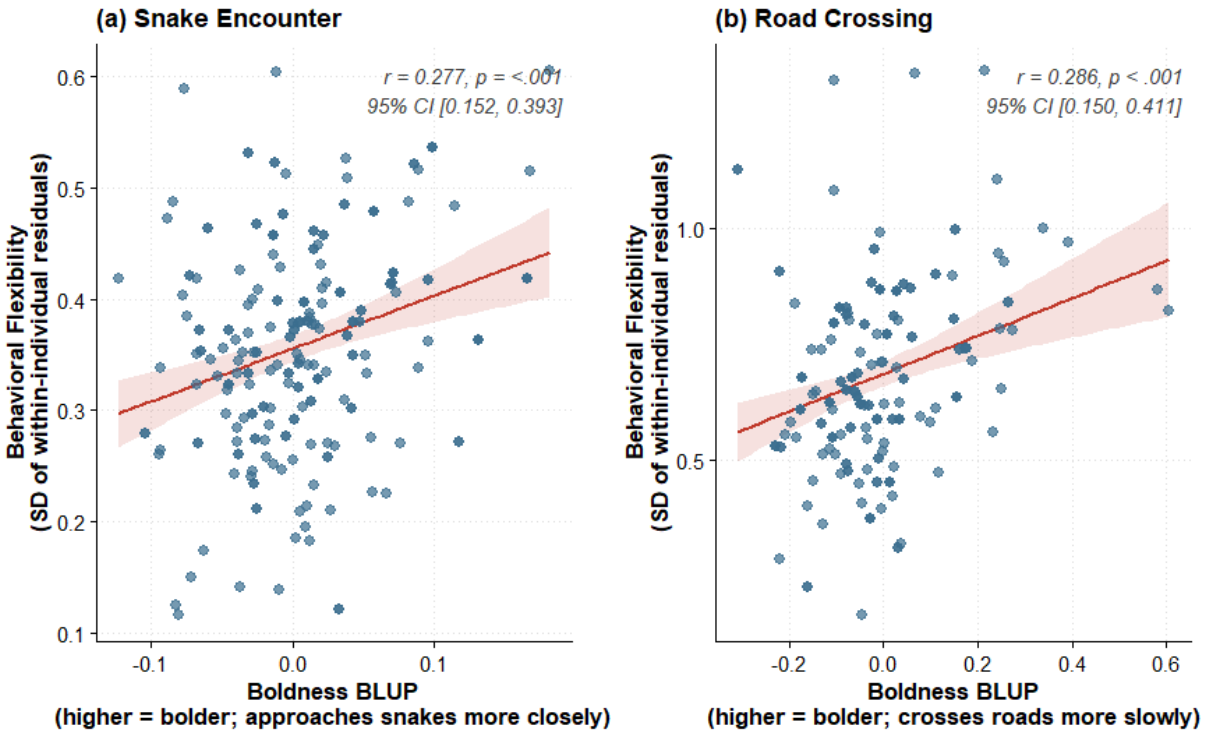


Figure 2.5: Positive relationship between boldness and behavioral flexibility in (a) snake encounter and (b) road crossing contexts.

Cross-Context Boldness Consistency

Individual boldness BLUPs derived from the snake approach model showed no significant cross-context correlation with boldness BLUPs derived from the road-crossing model at the population level (Spearman's ρ = −0.083, $p = 0.565$, $n = 50$). Among adult and subadult males specifically, the cross-context correlation was similarly near-zero and non-significant (Spearman's ρ = 0.051, $p = 0.867$, $n = 14$; see Figure 2.6). These results indicate that an individual's relative boldness in a predator-relevant context does not predict their relative boldness in an anthropogenic risk context, providing evidence against a domain-general boldness construct in this population and consistent with the jingle fallacy concern articulated earlier. The

two assays appear to be measuring distinct context-specific behavioral tendencies rather than a unified latent trait.

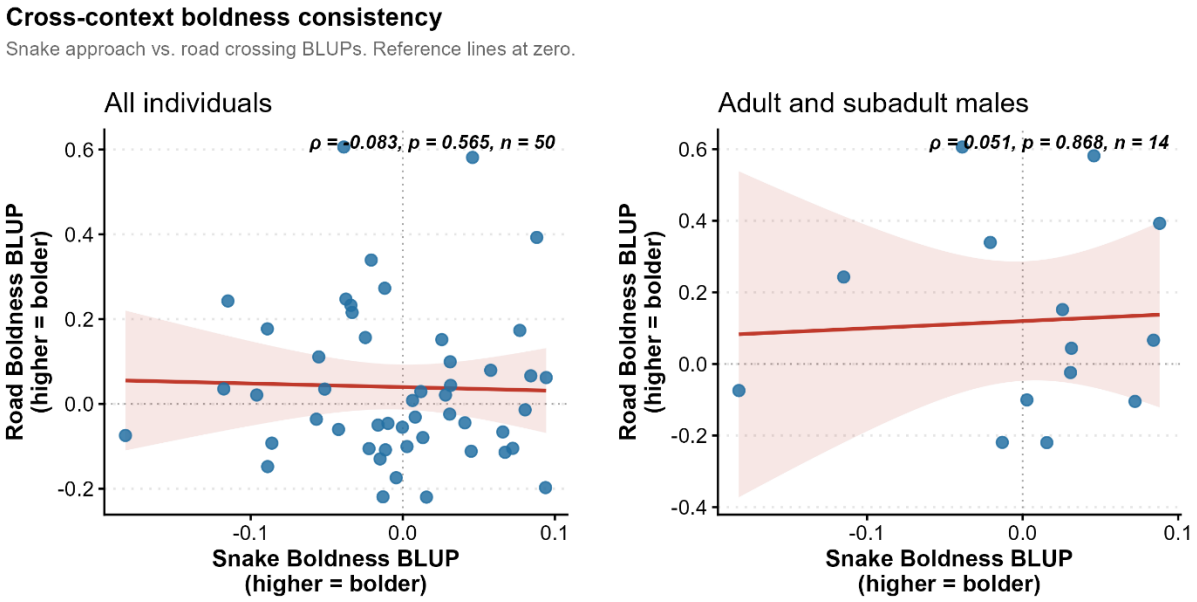


Figure 2.6: Absence of cross-context boldness consistency. Each point represents one individual with data in both behavioral contexts (Panel A: all individuals, N = 50; Panel B: adult and subadult males, N = 14). Reference lines drawn at zero on both axes (population mean after controlling for fixed effects).

Personality Rating Validity: Fearfulness

Contrary to the expectation that observer-rated fearfulness would negatively predict behavioral boldness, fearfulness ratings showed no significant relationship with individual boldness in either context. In the primary analyses (minimum five behavioral observations, minimum two raters), fearfulness was uncorrelated with road-crossing boldness (Spearman ρ = 0.039, 95% CI [−0.198, 0.271], $p = 0.749$, $n = 70$) and with snake-approach boldness (Spearman $\rho = 0.167$, 95% CI [−0.071, 0.387], $p = 0.246$, $n = 50$). A sensitivity analysis relaxing both the observation threshold (to a minimum of two) produced comparable null results : $r = 0.0004$ (95% CI [−0.235, 0.235], $p = 0.997$) for road boldness and $r = 0.167$ (95% CI [−0.071, 0.387], p

= *0.*246) for snake boldness. The null result therefore does not appear to be an artifact of sample size or threshold choices.

Aggressiveness ratings showed robust convergent validity across all three aggression measures, while fearfulness ratings showed no predictive validity for behavioral boldness is either context. Together these results suggest that the personality instrument captures genuine individual variation in aggressiveness but does not capture meaningful variation in the behavioral risk-taking tendencies indexed by the snake approach and road crossing assays. This may reflect limited behavioral visibility of fearfulness relative to aggressiveness, bold or shy behavior toward snakes and road crossings may be less salient to observers than direct agonism or may indicate that the fearfulness construct as rated by observers captures something functionally distinct from the context-specific boldness tendencies measured in this study.

Objective 3: Boldness–Aggression Behavioral Syndrome

Population-Level Syndrome Test: Random Slopes Model

To test whether aggression and snake boldness co-vary within individuals I fit random slopes mixed-effects models in which both a random intercept and a random slope for snake boldness were estimated for each individual, with aggression context (toward adult males, adult females, and non-adults) included as a fixed effect. Models were fit separately for adult and subadult males (primary analysis), adult females, and juveniles (comparative/exploratory).

At the population level, snake boldness did not significantly predict aggression rates in any demographic group (adult males: $\beta = -0.052$, SE = 0.067, $p = 0.498$; adult females: $\beta = -0.015$, SE = 0.075, $p = 0.846$; juveniles: $\beta = 0.013$, SE = 0.039, $p = 0.730$), providing no evidence for a consistent population-level boldness–aggression syndrome. This null result was

consistent across all three groups and converges with the Spearman rank correlation results reported above.

Despite the absence of a population-level effect, the adult male model revealed substantial individual heterogeneity in the strength and direction of the boldness–aggression relationship. The distribution of individual slope BLUPs showed considerable spread around zero, with some males exhibiting strong positive coupling between boldness and aggression, others showing negative coupling, and most clustering near zero - a pattern inconsistent with a simple, uniform syndrome but consistent with individual-level strategic variation in behavioral organization (Figure 2.7). Although the random slope variance estimate was technically bounded (intercept–slope correlation = 1.00, indicating a boundary solution), the spread of individual BLUPs is interpretable as evidence of meaningful individual heterogeneity. The positive intercept–slope correlation (though bounded) reflects the biologically coherent pattern that higher-aggression individuals tend to show stronger boldness–aggression coupling. The random slope term did not significantly improve model fit via likelihood ratio test (LRT: $\chi^2(2) = 0.095$, $p = 0.954$), indicating that individual heterogeneity in syndrome expression, while visually apparent, could not be formally distinguished from zero with the available sample size (N = 29 males).

The female and juvenile models both returned singular fits (both models having intercept-slope correlations of −1.00) with non-significant random slope LRTs (female: $\chi^2(2) = 2.64$, $p = 0.267$; juvenile: $\chi^2(2) = 0.097$, $p = 0.953$), precluding reliable estimation of individual heterogeneity in syndrome expression for these groups. These results should be interpreted with caution and are reported for comparative purposes only.

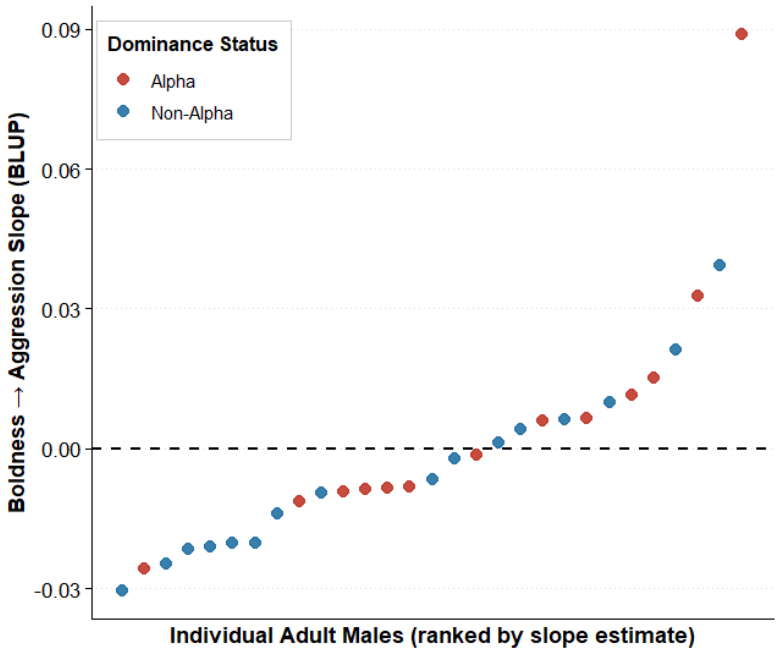


Figure 2.7. Individual variation in the strength and direction of the boldness aggression relationship in adult male capuchins. Points represent best linear unbiased predictors (BLUPs) of individual-specific slopes from a random slopes model (aggression ~ snake_boldness + aggression_context + (1 + snake_boldness | monkey_id); N = 29 males). Color indicates whether the individual held alpha status at any point during the observation period.

Population-Level Syndrome Test: Bivariate Correlations

Bivariate Spearman rank correlations between individual boldness BLUPs and logit-transformed aggression rates provide a consistent and convergent picture (Table 2.4). No significant boldness–aggression association was detected after FDR correction for any of the six boldness–target combinations. Snake boldness BLUPs were uncorrelated with aggression toward adult males ($\rho = -0.153$, adult females ($\rho = -0.314$), or non-adults ($\rho = -0.201$). Road boldness BLUPs were likewise uncorrelated with aggression toward adult males ($\rho = 0.017$) and non-adults ($\rho = -0.218$). The one nominally significant uncorrected association - road boldness and female-directed aggression ($\rho = 0.396$, raw $p = 0.037$) - did not survive FDR correction (adjusted

*p = 0.*222) and falls below the minimum detectable effect size at 80% power for this sample size (ρ ≥ 0.50 at *n* = 28) and is therefore interpreted as a chance finding. These results converge with the random slopes analysis in providing robust evidence against a consistent, uniform population-level boldness–aggression syndrome.

Table 2.4. Spearman rank correlations between individual boldness BLUPs and logit-transformed unprovoked aggression rates toward each social target class in adult and subadult males. p-values adjusted using FDR correction across all six comparisons within each boldness context.

Boldness context	**Aggression target**	**N**	**ρ**	**p (raw)**	**p (FDR)**
Snake Approach	→ Adult males	29	−0.153	.426	1.000
	→ Adult females	29	−0.314	.097	.580
	→ Non-adults	29	−0.201	.296	1.000
Road Crossing	→ Adult males	28	0.017	.933	1.000
	→ Adult females	28	0.396	.037	.222
	→ Non-adults	28	−0.218	.265	1.000

Note. Minimum detectable effect size at 80% power, α = .05: ρ ≥ 0.50 (n = 28–29). FDR-adjusted p-values for all comparisons reported; adjusted values of 1.000 indicate the raw p-value multiplied by 6 exceeds 1.

Individual Heterogeneity in Syndrome Expression

Despite the absence of a population-level syndrome, examination of individual data revealed substantial heterogeneity in how boldness and aggression were coupled across individuals (Figure 2.7). Plotting each individual's snake boldness BLUP against their logit-transformed rate of unprovoked male-directed aggression and partitioning individuals into quadrants by median splits of both variables showed approximately equal distribution across all four quadrants: bold–aggressive (*n* = 6), bold–unaggressive (*n* = 8), shy–aggressive (*n* = 8), and

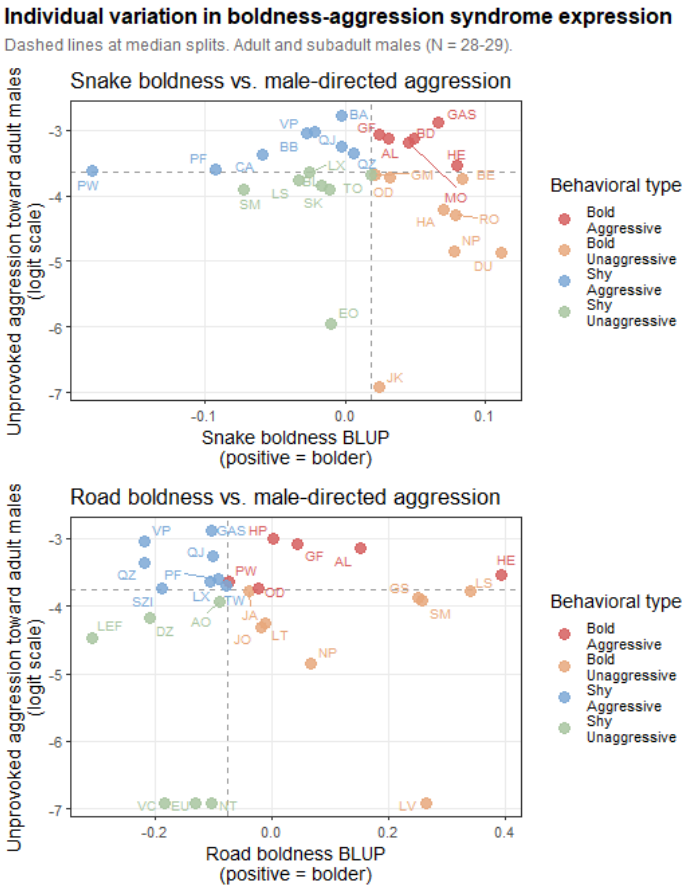


Figure 2.8 Individual variation in boldness-aggression syndrome expression.

shy–unaggressive (n = 7). An analogous plot substituting road-crossing boldness BLUPs produced a similarly even distribution across quadrants, reinforcing the conclusion that no consistent directional coupling between boldness and aggression is detectable regardless of the risk context used to index boldness.

The even distribution across quadrants is inconsistent with any uniform positive or negative syndrome and instead indicates that the boldness–aggression relationship varies substantially in direction and magnitude across individuals. An exploratory latent profile analysis on the expanded sample including both sexes (N = 112 individuals with snake boldness and aggression data) did not identify a discrete behavioral typology that would account for this heterogeneity. The four-profile solution suggested by BIC-based model selection (BIC – 1195.7) yielded a classification in which the overwhelming majority of individuals collapsed into a single profile with a mean posterior classification probability of 0.000, indicating that the model could not reliably distinguish this group from the population mean. No solution approached the conventional reliability threshold of 0.70 required to consider profile assignments meaningful. Furthermore, sex did not significantly predict profile membership ($\chi^2(2) = 2.88$, $p = 0.411$).

These results indicate that individual variation in syndrome expression is better characterized as a continuous distribution of coupling strengths than as membership in discrete behavioral types, a conclusion consistent with the visual impression of the random slope distribution in Figure 2.8 and with the continuous spread of individuals across all four quadrants of the boldness–aggression space.

Discussion

This study set out to characterize the structure of behavioral variation in wild white-faced capuchins by examining the repeatability of boldness and aggression as distinct traits, their covariance across contexts, and the validity of psychometric personality ratings as proxies for observed behavior. The results present a coherent, if unexpected, picture that (a) these behaviors are measurably consistent within individuals but far more labile than the cross-taxa literature would predict, (b) they do not organize into a uniform population-level boldness–aggression syndrome, and (c) their coupling varies substantially across individuals in ways that cannot be attributed to a simple typology. Together, these findings challenge the universality of the syndrome framework as commonly applied and point toward a more nuanced, individual-heterogeneity-centered view of behavioral integration in a cognitively complex primate.

Aggression is Target-Specific, Not Trait-Like

The most structurally important finding from the aggression analyses is not that repeatability was low, though it was, but that it was *differentially* low in a pattern consistent with functional target-specificity rather than random noise. Female-directed aggression was the most repeatable form (R = 0.074 full, 0.087 males-only), non-adult-directed aggression the least (R = 0.036 full), and male-directed aggression intermediate (R = 0.048 full). The relatively stronger association with female-directed aggression may reflect the particular political salience of male-

female relationships in this population. Securing female alliance is likely a prerequisite for alpha male tenure, and alpha males invest more heavily in affiliative and coalitionary relationships with high-ranking females than with other males (Perry, 1997, 1998). Female-directed aggression may therefore reflect more stable, individually characteristic patterns of relationship investment than male-direct competitive aggression, which fluctuates with rank dynamics and challenger availability. Non-adult-directed aggression showed the lowest repeatability ($R = 0.036$ full), consistent with its status as the least competitively consequential target class. That aggression repeatability varies systematically across target classes – highest where the social stakes are greatest, lowest where they are least – supports the interpretation that aggression in this population is regulated strategically at the level of the individual rather than emitted as a uniform expression of a generalized "aggressiveness" disposition.

The lack of a significant correlation between male-directed and female-directed aggression rates ($\rho = -0.114$, 95% CI [−0.365, 0.152], $p = 0.372$) in adult males further supports the interpretation that aggression toward different social targets does not reflect a single, generalized aggressiveness trait. Male-directed aggression was also unrelated to aggression toward non-adults ($\rho = 0.116$, 95% CI [−0.151, 0.366], $p = 0.367$), whereas female-directed and non-adult-directed aggression were modestly positively correlated ($\rho = 0.333$, 95% CI [0.078, 0.547], $p = 0.008$). Thus, rather than indicating a trade-off in which males that are highly aggressive toward rivals are less aggressive toward females, these results suggest that male-directed aggression varies largely independently of aggression toward other social targets. This target specificity is consistent with the ecology of male capuchins, in which male-directed aggression is primarily competitive, carrying high costs and targeting rivals for rank and reproductive access, while aggression towards females occurs in a different social and political

context. Males in this population invest in relationships with resident females both by supporting females in coalitions against other females and by defending females against male aggression, with female political support in return representing a critical resource for alpha male tenure (Fedigan et al., 2021; Perry, 1997, 1998). The differing social functions, costs, and potential payoffs of aggression across target classes may therefore favor context-dependent regulation rather than the consistent expression of a unitary aggressiveness trait.

These results place white-faced capuchins in the company of a growing set of taxa in which aggression has been found to partition by social context rather than load on a single trait axis. Bell (2005) demonstrated target-specificity in threespine sticklebacks, where aggression toward predators and toward conspecifics constitute partially independent behavioral modules. In primates, the social complexity hypothesis (Dunbar, 1998) predicts precisely this kind of behavioral modularity - species living in large, differentiated social networks benefit from fine-grained, context-sensitive behavioral regulation rather than rigid, domain-general traits. The present findings are consistent with this expectation and suggest that capuchins are a particularly clear case where the assumption of a single aggressiveness factor, embedded in much of the personality rating and behavioral syndrome literature, may obscure more than it reveals.

Boldness Repeatability: Context-Dependency and the Behavioral Flexibility Paradox

Low Repeatability in Both Contexts

Repeatability was strikingly low in both risk contexts: $R = 0.023$ for snake-approach boldness and $R = 0.070$ for road-crossing boldness after log transformation. Both values fall well below the cross-taxa meta-analytic mean of approximately 0.37 reported by (Bell et al., 2009) and toward the lower tail of the distribution for single-context behavioral traits in wild vertebrates.

The variance decomposition from the snake-approach model is especially informative - 44% of variance in log approach distance was attributable to between-encounter differences, compared to only 1.7% to stable individual differences. The specific characteristics of each snake encounter collectively explain roughly 26 times more of the behavioral variation than individual identity does. This does not mean that individuals are interchangeable, but that the same individual reliably adjusts its behavior to the state of the environment, and that this adjustment overwhelms whatever consistent individual tendency is detectable in the data. From this perspective, the low repeatability is less likely to reflect a measurement failure and more likely means that snake-approach behavior in capuchins is a highly context-sensitive response rather than a stable personality trait.

The road-crossing context tells a somewhat different story. With only one random-effect grouping level (no encounter-level nesting), all residual variance that was absorbed by snake_event in the snake model is distributed between individual identity and within-individual residual. The resulting $R = 0.070$ is higher than the snake estimate, but this difference may partly reflect the absence of an encounter-level random effect rather than a genuine difference in the stability of the underlying trait. What both contexts share is a dominant role for situational factors (encounter characteristics in the snake case, car presence, and age-related context in the road case) over individual consistency.

The Boldness–Flexibility Association

The finding that bolder individuals show *greater* within-individual variability in their behavior replicated across both contexts (snake: $r = 0.235$; road: $r = 0.325$) and inverts the prediction of classical proactive–reactive coping style theory. In the standard framework (Coppens et al., 2010; Koolhaas et al., 1999), the proactive coping style is characterized by bold,

aggressive behavior *and* by behavioral rigidity. Proactive individuals apply routine solutions, exploit familiar environments, and resist adapting to novelty. If boldness in capuchins indexed a proactive coping style in this sense, bolder individuals should show *less* variability, and tighter, more stereotyped responses than shy individuals.

Instead, we observe the reverse pattern. Bold individuals explore a wider range of responses around their own expected behavior, suggesting that boldness in this species facilitates rather than forecloses behavioral sampling. This is consistent with an interpretation in which boldness functions as a prerequisite for environmental exploration, what might be called an *exploration-enabling* rather than a *routine-maintaining* phenotype. White-faced capuchins are extractive foragers with well-documented innovations and social traditions (Fragaszy et al., 2004; S. Perry, 2011). For a species whose ecological success depends heavily on behavioral invention and the modification of acquired behaviors, bold engagement with novel or risky situations may be the mechanism that generates the behavioral variation from which learned solutions are selected. In this light, the boldness–flexibility association is not anomalous but ecologically coherent. Boldness enables the sampling that produces flexibility.

The capuchin case thus adds to a small but growing literature suggesting that the proactive–reactive axis, as characterized in the rodent and fish literature, may not translate cleanly to taxa with greater cognitive complexity and behavioral plasticity. Groothuis & Trillmich (2011) have argued that the rigidity associated with proactive coping styles in laboratory rodents may itself be a phenotype calibrated to relatively impoverished, predictable environments rather than a universal property of bold behavioral types. In more ecologically complex, socially dynamic environments, like those inhabited by a long-lived, group-living

primate, selection may favor bold individuals who can also flexibly adjust, rather than those who can only respond reliably.

Cross-Context Independence

Snake-approach boldness and road-crossing boldness were uncorrelated at the population level (Spearman's $\rho = -0.067$) and within adult and subadult males specifically ($\rho = 0.055$). This null cross-context result is structurally related to, and helps explain, the low within-context repeatability. If the same "boldness" construct drove both behaviors, one would expect not only within-context consistency but between-context consistency. The absence of both supports the conclusion that these two risk-taking measures are tapping distinct behavioral systems rather than a domain-general boldness trait.

This finding is consistent with the jingle fallacy concern articulated in the introduction. Labeling both behaviors "boldness" because they both involve approaching something risky does not mean they share a common cause. Snake encounters and road crossings differ in their sensory characteristics, their evolutionary history as sources of mortality risk, their social context, and in the motivational and cognitive demands they impose. That individuals do not show consistent relative tendencies across these two situations is the expected result if behavior is proximately controlled by context-sensitive decision processes rather than by a single, stable risk-tolerance parameter.

The broader implication is that studies that use a single behavioral assay and label the resulting individual differences "boldness" are likely measuring something more specific than they are claiming. Convergent validity across contexts, or its explicit acknowledgment as absent, should be a minimum standard for personality research in field settings.

Personality Ratings

The personality rating results present a clean dissociation that is both practically useful and theoretically informative. Aggressiveness ratings showed robust convergent validity across all three behavioral aggression measures - the associations with female-directed aggression ($\beta = 0.320$) and with male-directed and non-adult-directed aggression ($\beta \sim 0.19$) were all significant and meaningfully sized, indicating that long-term observers are genuinely detecting individual differences in agonistic behavior when they assign aggressiveness scores. The gradient in effect sizes is itself informative. Conspicuous, repetitive female-directed agonism may be among the most perceptually available behavioral signals to observers who spend months or years with individual animals, making it the primary empirical basis for their aggressiveness judgments.

Fearfulness ratings, by contrast, showed essentially zero predictive validity for behavioral boldness in either context. Effect sizes were negligible ($r \sim 0.07–0.11$) and confidence intervals were wide, consistent with true near-zero relationships rather than merely imprecise estimates of modest effects. The sensitivity analysis, which relaxed both the observation and rater thresholds, reproduced these null results with somewhat tighter intervals, ruling out threshold choice as an explanation.

Several non-mutually exclusive accounts could explain this dissociation. First, the behavioral expression of boldness in snake and road contexts may simply be less observable to raters than aggression. Aggressive interactions are public, often dramatic, and directly affect the rater's own social environment. Snake encounter and road crossing protocols were inconsistently applied over the study period, and many raters were likely not cued to be attentive to behavior in these contexts if not part of these particular data collection efforts. If raters are making fearfulness attributions based on a different behavioral substrate like for example, generalized wariness or fear of conspecifics, then the rating may be a valid index of a somewhat different

construct than the boldness measured here. Second, the fearfulness facet from the Manson and Perry (2013) instrument may reflect emotional reactivity or neophobia rather than the risk-tolerance dimension indexed by approach-avoidance responses to ecologically specific stimuli. These are conceptually related but not identical constructs, and their behavioral correlates need not be the same.

The practical implication is that within this population, aggressiveness ratings can serve as useful proxies for observed agonistic behavior and can be incorporated into models where direct behavioral data are unavailable. Fearfulness ratings, at least as predictors of risk-taking in antipredator and anthropogenic contexts, cannot.

No Population-Level Syndrome, But Meaningful Individual Heterogeneity

The Null Population-Level Result

No evidence emerged for a consistent population-level boldness–aggression syndrome. The random slopes model found no average association between boldness and aggression across individuals ($\beta = 7.53 \times 10^{-5}$, $p = 0.418$), and the bivariate Spearman correlations were uniformly non-significant after FDR correction. These two approaches converge on the same conclusion from different analytical angles. The random slopes model asks whether boldness predicts aggression as a fixed population trend, while the bivariate correlations ask whether individuals who rank higher on boldness also rank higher on aggression, and both find no signal.

The null result is not obviously attributable to power limitations, though low power must be acknowledged. The bivariate analyses had N = 28–29 with minimum detectable effects at 80% power of $\rho \geq 0.50$. The absence of even nominally suggestive patterns in the expected direction across six independent correlation tests - with effect sizes ranging from $\rho = -0.314$ to $\rho = 0.396$ and centering near zero - is more consistent with a true null than with a real moderate-

sized effect being masked by noise. The direction of the associations is also not consistently positive. Several correlations are negative, which would be unexpected if a moderate positive syndrome existed but was merely underpowered.

How does this result compare to the existing literature? The boldness–aggression syndrome has been documented in numerous taxa, from sticklebacks (Bell, 2005) to great tits (Dingemanse & Dochtermann, 2013). Meta-analytic work suggests that the syndrome is reliably found across taxa and relatively robust to methodological variation (Bell, 2005). The absence of a detectable population-level syndrome in capuchins is therefore noteworthy. One interpretation is taxonomic, that the syndrome literature is disproportionately based on species with fast life histories, high reproductive rates, and relatively simple social structures, and it is plausible that the syndrome, if it exists at all in slow-lived, cognitively complex primates, is expressed differently or not expressed as a fixed population-level correlation in such taxa. A second interpretation is socioecological: in a species where aggression must be precisely calibrated to target identity, rank dynamics, and coalition politics, a rigid coupling of boldness to general aggressiveness would be maladaptive. A male who is bold in predator contexts but strategically discriminating regarding his targets of aggression would likely outcompete a male whose boldness spills over uniformly into aggression.

Individual Heterogeneity as the Substantive Finding

The more theoretically interesting result is not the null fixed effect but the significant variance in random slopes. The strength of the boldness–aggression relationship varies meaningfully across individuals, with some showing strong positive coupling, most near-zero coupling, and a minority showing negative coupling. The quadrant visualization makes this more concrete: bold–aggressive, bold–unaggressive, shy–aggressive, and shy–unaggressive

phenotypes all exist in roughly equal numbers. There is no modal capuchin "type" that represents the population tendency; the population is heterogenous.

This finding connects to a growing body of work arguing that between-individual variance in syndrome expression, rather than the population-level mean correlation, is itself an ecologically and evolutionarily important quantity (Dingemanse & Dochtermann, 2013; Westneat et al., 2015). (Royauté & Dochtermann, 2017) demonstrated in field crickets that developmental experience (specifically early-life crowding stress) increased among-individual variance in syndrome expression, meaning that adversity produced not a uniform shift in phenotype but a diversification of individual coupling strengths. The between-individual variance in coupling observed here may reflect analogous processes. Individuals may differ in the degree to which their risk-taking and agonistic systems are developmentally integrated, and this heterogeneity may itself be maintained by frequency-dependent selection or by the diversity of social niches available within capuchin groups (Bergmüller & Taborsky, 2010; Laskowski et al., 2022).

The failure of the latent profile analysis (mean posterior classification probability = 0.33 for the three-profile solution) to identify discrete behavioral types reinforces the conclusion that this heterogeneity is continuous rather than categorical. The boldness–aggression space in this population appears to be occupied by a broad, relatively uniform cloud of individuals rather than by distinguishable clusters.

Implications for the POLS Framework and Primate Personality Research

The Pace-of-Life Syndrome framework predicts that bold, aggressive, risk-prone phenotypes should co-occur as part of an integrated "fast" life-history strategy (Biro & Stamps, 2010; Réale et al., 2010). The present results do not support this prediction for white-faced

capuchins, at least not as a population-level phenomenon. If anything, the target-specificity of aggression, the lack of cross-context boldness consistency, and the near-zero population-level boldness–aggression correlation suggest that the behavioral components of the POLS fast end are not integrated into a package in this species but are instead regulated independently in response to context.

This may reflect the particular demands of capuchin social life. The POLS framework was developed primarily with data from species where the "fast–slow" axis captures straightforward trade-offs between current reproduction and survival. In a long-lived, group-living primate where male reproductive success depends on achieving and maintaining alpha status through a complex combination of competitive aggression, coalition formation, male–male cooperation, and female alliance, the optimal behavioral strategy is almost certainly more conditional and context-sensitive than a fixed fast/slow type. An adult male capuchin who has just achieved alpha status faces a quite different optimization problem from one who is newly immigrated or who is a peripheral adult - the first should be relatively risk-averse (asset protection), the second should be relatively bold and competitively aggressive. And the optimal combination of boldness and aggression will depend not just on the individual's own phenotype but on the phenotypes of his competitors and allies. Static, population-level syndrome correlations capture none of this dynamic.

More broadly, these results suggest that the theoretical framework connecting behavioral syndromes to life-history variation in primates may need to be substantially modified before it can make accurate predictions. The expectation that behavioral traits will covary in fixed patterns because they share physiological substrates or developmental pathways - the mechanistic logic underlying POLS (Careau et al., 2008) - may be less applicable in taxa where the neuroendocrine

and cognitive systems supporting behavior are sufficiently complex and plastic to allow context-dependent dissociation of traits that are coupled in simpler organisms.

Limitations and Future Directions

The syndrome analyses were constrained to N = 28–29 adult and subadult males with both boldness and aggression data - a sample size that provides adequate power to detect large effects ($\rho \geq 0.50$) but limited power to detect moderate effects ($\rho \sim 0.30$). The null population-level results are therefore interpretable as evidence against large uniform syndromes, but they cannot rule out moderate effects. Future analyses should prioritize expanding the boldness dataset to include a larger share of females and age classes to permit full sample syndrome tests and should consider whether the constraint of requiring both behavioral and boldness data for the same individual can be relaxed through model-based approaches that integrate across incomplete observation records.

The within-context repeatability estimates, particularly for snake boldness, should be interpreted with the understanding that these are adjusted repeatabilities from models with fixed effects for age class and snake species; they represent the residual individual consistency after removing these substantial sources of variation. Naive (unadjusted) repeatability would be higher. The theoretical question of whether the relevant measure for syndrome analysis is the adjusted or unadjusted individual difference is not trivial. If boldness is partly a developmental state that changes with age, the adjusted BLUP captures the deviation from expected age-class behavior, which may be the most relevant individual-difference variable for predicting syndrome coupling. Alternatively, the total between-individual variance including age-class differences may be more relevant if the goal is to understand population-level behavioral variation broadly.

A related methodological consideration is whether the unusually low repeatability estimates reported here are partly a consequence of having far more observations per individual than is typical of studies included in the Bell et al. (2009) meta-analysis. The typical study contributing to the Bell et al. (2009) meta-analytic benchmark used approximately two repeated measures per individual (Wolak et al., 2012), which may constrain within-individual variance estimation and yield upwardly biased repeatability. The present dataset, by contrast, includes a mean of 26.6 snake encounter observations and 50.5 road crossing observations per individual, providing a much more thorough sampling of within-individual behavioral variation - including variation attributable to context-specific adjustment that would be invisible in sparser designs. A resampling simulation could directly test whether this discrepancy explains part of the gap between our estimates and the meta-analytic benchmark. Such a simulation would repeatedly draw small subsets of observations per individual (e.g., $n = 2$), matched as closely as possible on contextual factors known to influence behavior (snake species, car presence, and temporal proximity of observations), re-estimate repeatability from each draw, and compare the resulting distribution against the full-data estimates. If repeatability estimates from subsampled data converge on values closer to the meta-analytic mean, this would suggest that the low repeatability in the present study reflects the greater statistical resolution afforded by dense longitudinal sampling. This simulation is planned as a supplementary analysis prior to publication.

The fearfulness rating null cannot resolve whether fearfulness ratings fail to predict behavioral boldness (a) because the ratings are invalid, (b) because they capture a genuinely distinct construct (neophobia, emotional reactivity) that is only loosely linked to approach-avoidance boldness, or (c) because the behavioral measures used here do not sample the

behavioral domain that observers use when assigning fearfulness scores. The most important next step is the one that this chapter is explicitly designed to enable: testing whether individual heterogeneity in syndrome expression - the variance in random slopes that emerged as the central finding - is predicted by early-life adversity. The developmental canalization framework (Royauté & Dochtermann, 2017) predicts that individuals who experienced greater social instability during development should show either tighter syndrome integration (if adversity channels individuals into a "committed" proactive strategy) or greater syndrome variance (if adversity disrupts normative development and produces a wider range of outcomes). Chapter 4 attempts to addresses this question directly.

Conclusions

This study provides a comprehensive characterization of boldness and aggression as behavioral traits in wild white-faced capuchins, and in doing so raises important questions about the generalizability of behavioral syndrome theory to cognitively complex, socially dynamic primates. The main findings are as follows. Aggression in adult male capuchins is target-specific, not trait-like, and it is moderately consistent within target class but negatively correlated across the male- and female-directed targets, suggesting independent motivational regulation rather than a general aggressiveness factor. Boldness in both predator and anthropogenic risk contexts is highly labile, dominated by encounter-specific variance, and uncorrelated across contexts, indicating that "boldness" as assessed in these two settings measures context-sensitive risk management rather than a stable individual trait. Bolder individuals in both contexts show greater behavioral flexibility, inverting the prediction of proactive coping style theory and suggesting that boldness in a cognitively complex primate may function as an exploration enabler. Aggressiveness personality ratings have genuine convergent validity with observed

agonistic behavior, but fearfulness ratings do not predict behavioral boldness in either context. And the boldness–aggression syndrome, while absent at the population level, is present as a meaningful dimension of *individual* heterogeneity. Individuals vary substantially in the degree and direction of their boldness–aggression coupling.

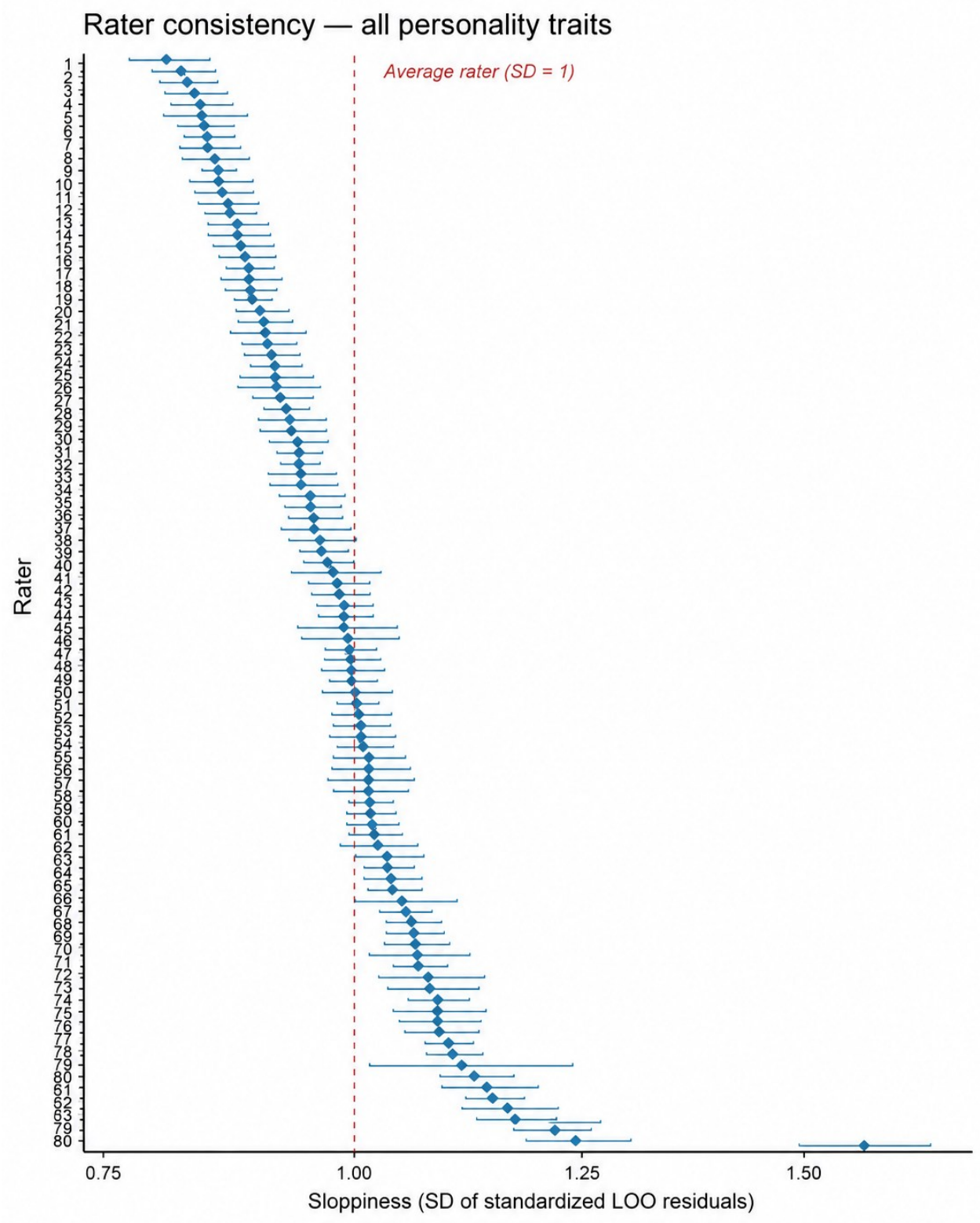
These findings collectively suggest that capuchins maintain a degree of behavioral independence between risk-taking and agonism that is greater than that documented in most syndromic species, and that this independence may be adaptive given the precision with which these animals must calibrate aggression to social target and context. The population does not present a unified behavioral architecture, but instead a distribution of architectures, and that distribution is the phenomenon that requires explanation. Whether early-life adversity is one of the processes that generates this distribution is the question to which this dissertation now turns.

Supplementary Materials

Leave-one-out (LOO) Rater reliability assessment.

For each rating, a LOO peer mean was computed as the average score assigned by all other raters to the same individual on the same trait during the same annual observation period; only monkey-age-trait groups with at least two distinct raters were included. The residual between each rater's score and the LOO peer mean was standardized within each trait by dividing by the global standard deviation of residuals for that trait, placing all traits on a common scale. Each rater was then characterized by a *sloppiness* score defined as the standard deviation of their standardized residuals across all ratings and thus reflecting how erratically they deviated from peer consensus. Asymmetric 95% confidence intervals for sloppiness were calculated using the chi-square distribution. Raters whose lower 95% confidence bound exceeded 1.3, the maximum lower bound observed among the remaining rater pool, were flagged as outliers. One rater met this criterion (lower CI = 1.5), with a sloppiness score substantially elevated above the bulk distribution (all other raters' lower bounds ≤ 1.3). 783 ratings (out of 246,837 or 0.3%) attributed to this rater were excluded from subsequent analyses. The threshold reflects a natural discontinuity in the sloppiness distribution rather than an a priori cutoff, and results were robust to alternative exclusion criteria (analyses including this rater have been retained and show no substantive change in results, see below).

Supplementary Figure S2.1: Rater consistency across all personality trait items Each point represents one rater's 'sloppiness' score, horizontal bars show asymmetric 95% CI calculated using chi-square distribution.



Supplementary Table S2.1. Personality rating validity results prior to "sloppy" rater exclusion.
Analyses were initially conducted using mean trait ratings computed from all available raters. One rater was subsequently identified as an outlier via leave-one-out peer disagreement analysis (sloppiness score substantially elevated above the rater pool; lower 95% CI bound = 1.5, exceeding the natural gap in the distribution at 1.3) and excluded from primary analyses. Results reported in the main text reflect the rater-filtered dataset. Pre-exclusion results are reported below for comparison; substantive conclusions were unchanged by the exclusion.

S1a. Aggressiveness rating validity (pre-exclusion)
Inter-rater reliability: ICC(3,k) = 0.99.

Outcome	β	SE	z	p	95% CI
Unprovoked → adult males	0.186	0.048	3.83	< 0.001	[0.091, 0.281]
Unprovoked → adult females	0.320	0.065	4.95	< 0.001	[0.193, 0.447]
Unprovoked → non-adults	0.192	0.043	4.43	< 0.001	[0.107, 0.277]

Note. β = log-odds coefficient for a one SD increase in aggressiveness rating. N = 370 observation periods, 285 individuals. All models fit using bobyqa optimizer with age class as a fixed covariate and individual monkey identity as a random intercept.

S1b. Fearfulness rating validity (pre-exclusion)
Inter-rater reliability: ICC(3,k) = 0.96

Context	N	r	95% CI	p	Analysis
Road crossing boldness	70	−0.013	[−0.247, 0.223]	0.916	Primary
Snake approach boldness	50	0.137	[−0.157, 0.408]	0.360	Primary
Road crossing boldness	83	-0.022	[-0.237, 0.195]	0.843	Sensitivity
Snake approach boldness	50	0.165	[-0.119, 0.424]	0.252	Sensitivity

Note. Spearman rank correlations between mean fearfulness rating and individual boldness BLUPs.

References

Altmann, J. (1974). *Observational Study of Behavior: Sampling Methods*. https://doi.org/10.1163/156853974X00534

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R. A., Gluckman, P., Godfrey, K., Kirkwood, T., Lahr, M. M., McNamara, J., Metcalfe, N. B., Monaghan, P., Spencer, H. G., & Sultan, S. E. (2004). Developmental plasticity and human health. *Nature*, *430*(6998), 419–421. https://doi.org/10.1038/nature02725

Baugh, A. T., Oers, K. van, Dingemanse, N. J., & Hau, M. (2014). Baseline and stress-induced glucocorticoid concentrations are not repeatable but covary within individual great tits (*Parus major*). *General and Comparative Endocrinology*, *208*, 154–163. https://doi.org/10.1016/j.ygcen.2014.08.014

Beehner, J. C., Alfaro, J., Allen, C., Benítez, M. E., Bergman, T. J., Buehler, M. S., Carrera, S. C., Chester, E. M., Deschner, T., Fuentes, A., Gault, C. M., Godoy, I., Jack, K. M., Kim, J. D., Kolinski, L., Kulick, N. K., Losch, T., Ordoñez, J. C., Perry, S. E., … Wasserman, M. D. (2022). Using an on-site laboratory for fecal steroid analysis in wild white-faced capuchins. *General and Comparative Endocrinology*, *329*, 114109. https://doi.org/10.1016/j.ygcen.2022.114109

Bell, A. M. (2005). Behavioural differences between individuals and two populations of stickleback (*Gasterosteus aculeatus*). *Journal of Evolutionary Biology*, *18*(2), 464–473. https://doi.org/10.1111/j.1420-9101.2004.00817.x

Bell, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1611), 755–761. https://doi.org/10.1098/rspb.2006.0199

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

Bergmüller, R., & Taborsky, M. (2010). Animal personality due to social niche specialisation. *Trends in Ecology & Evolution*, *25*(9), 504–511. https://doi.org/10.1016/j.tree.2010.06.012

Biro, P. A., & Stamps, J. A. (2010). Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends in Ecology & Evolution*, *25*(11), 653–659. https://doi.org/10.1016/j.tree.2010.08.003

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, 9th International Symposium on Avian Endocrinology*, *163*(1), 208–213. https://doi.org/10.1016/j.ygcen.2008.12.013

Busch, D. S., & Hayward, L. S. (2009). Stress in a conservation context: A discussion of glucocorticoid actions and how levels change with conservation-relevant variables. *Biological Conservation*, *142*(12), 2844–2853. https://doi.org/10.1016/j.biocon.2009.08.013

Careau, V., Thomas, D., Humphries, M. M., & Réale, D. (2008). Energy metabolism and animal personality. *Oikos*, *117*(5), 641–653. https://doi.org/10.1111/j.0030-1299.2008.16513.x

Carrera, S. C., Godoy, I., Gault, C. M., Mensing, A., Damm, J., Perry, S. E., & Beehner, J. C. (2025). Stress responsiveness in a wild primate predicts survival across an extreme El Niño drought. *Science Advances*, *11*(4), eadq5020. https://doi.org/10.1126/sciadv.adq5020

Champagne, F. A., & Meaney, M. J. (2006). Stress During Gestation Alters Postpartum Maternal Care and the Development of the Offspring in a Rodent Model. *Biological Psychiatry*, *59*(12), 1227–1235. https://doi.org/10.1016/j.biopsych.2005.10.016

Charnov, E. L., & Berrigan, D. (1993). Why do female primates have such long lifespans and so few babies? *Or* Life in the slow lane. *Evolutionary Anthropology: Issues, News, and Reviews*, *1*(6), 191–194. https://doi.org/10.1002/evan.1360010604

Chase, I. D., Bartolomeo, C., & Dugatkin, L. A. (1994). Aggressive interactions and inter-contest interval: How long do winners keep winning? *Animal Behaviour*, 393–400.

Cockrem, J. F. (2013). Corticosterone responses and personality in birds: Individual variation and the ability to cope with environmental changes due to climate change. *General and Comparative Endocrinology, 10th International Symposium on Avian Endocrinology*, *190*, 156–163. https://doi.org/10.1016/j.ygcen.2013.02.021

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 B: Biological Sciences*, *365*(1560), 4021–4028. https://doi.org/10.1098/rstb.2010.0217

De Jong, G., & Gavrilets, S. (2000). Maintenance of genetic variation in phenotypic plasticity: The role of environmental variation. *Genetical Research*, *76*(3), 295–304. https://doi.org/10.1017/S0016672300004729

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The Adaptive Calibration Model of stress responsivity. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1562–1592. https://doi.org/10.1016/j.neubiorev.2010.11.007

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse, N. J., Edelaar, P., & Kempenaers, B. (2010). Why is there variation in baseline glucocorticoid levels? *Trends in Ecology & Evolution*, *25*(5), 261–262. https://doi.org/10.1016/j.tree.2010.01.008

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5%253C178::AID-EVAN5%253E3.0.CO;2-8

Ellis, B. J., & Del Giudice, M. (2014). Beyond allostatic load: Rethinking the role of stress in regulating human development. *Development and Psychopathology*, *26*(1), 1–20. https://doi.org/10.1017/S0954579413000849

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*(Suppl 1).

Fedigan, L. M., Hogan, J. D., Campos, F. A., Kalbitzer, U., & Jack, K. M. (2021). Costs of male infanticide for female capuchins: When does an adaptive male reproductive strategy become costly for females and detrimental to population viability? *American Journal of Physical Anthropology*, *176*(3), 349–360. https://doi.org/10.1002/ajpa.24354

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The Complete Capuchin: The Biology of the Genus Cebus*. Cambridge University Press.

Frankenhuis, W. E., Nettle, D., & Dall, S. R. X. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 20180110. https://doi.org/10.1098/rstb.2018.0110

Gholib, G., Heistermann, M., Agil, M., Supriatna, I., Purwantara, B., Nugraha, T. P., & Engelhardt, A. (2018). Comparison of fecal preservation and extraction methods for steroid hormone metabolite analysis in wild crested macaques. *Primates*, *59*(3), 281–292. https://doi.org/10.1007/s10329-018-0653-z

Groothuis, T. G. G., & Trillmich, F. (2011). Unfolding personalities: The importance of studying ontogeny. *Developmental Psychobiology*, *53*(6), 641–655. https://doi.org/10.1002/dev.20574

Gros-Louis, J., Perry, S., & Manson, J. H. (2003). Violent coalitionary attacks and intraspecific killing in wild white-faced capuchin monkeys (Cebus capucinus). *Primates*, *44*, 341–346.

Harvey, P. H., & Clutton-Brock, T. H. (1985). Life History Variation in Primates. *Evolution*, *39*(3), 559–581. https://doi.org/10.1111/j.1558-5646.1985.tb00395.x

Hau, M., Casagrande, S., Ouyang, J. Q., & Baugh, A. T. (2016). Glucocorticoid-Mediated Phenotypes in Vertebrates. In *Advances in the Study of Behavior* (Vol. 48, pp. 41–115). Elsevier. https://doi.org/10.1016/bs.asb.2016.01.002

Hau, M., Deimel, C., & Moiron, M. (2022). Great tits differ in glucocorticoid plasticity in response to spring temperature. *Proceedings of the Royal Society B: Biological Sciences*, *289*.

Hau, M., & Goymann, W. (2015). Endocrine mechanisms, behavioral phenotypes and plasticity: Known relationships and open questions. *Frontiers in Zoology*, *12*(Suppl 1), S7. https://doi.org/10.1186/1742-9994-12-S1-S7

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Isler, K., & Van Schaik, C. P. (2014). How humans evolved large brains: Comparative evidence. *Evolutionary Anthropology: Issues, News, and Reviews*, *23*(2), 65–75. https://doi.org/10.1002/evan.21403

Jack, K. M., & Fedigan, L. M. (2006). Why Be Alpha Male? Dominance and Reproductive Success in Wild White-Faced Capuchins (Cebus capucinus). In A. Estrada, P. A. Garber, M. S. M. Pavelka, & L. Luecke (Eds.), *New Perspectives in the Study of Mesoamerican Primates* (pp. 367–386). Kluwer Academic Publishers. https://doi.org/10.1007/0-387-25872-8_18

Jack, K. M., Schoof, V. A. M., Sheller, C. R., Rich, C. I., Klingelhofer, P. P., Ziegler, T. E., & Fedigan, L. (2014). Hormonal correlates of male life history stages in wild white-faced capuchin monkeys (*Cebus capucinus*). *General and Comparative Endocrinology*, *195*, 58–67. https://doi.org/10.1016/j.ygcen.2013.10.010

Kajokaite, K., Whalen, A., Panchanathan, K., & Perry, S. (2019). White-faced capuchin monkeys use both rank and relationship quality to recruit allies. *Animal Behaviour*, *154*, 161–169. https://doi.org/10.1016/j.anbehav.2019.06.008

Keller-Wood, M. E., & Dallman, M. F. (1984). Corticosteroid Inhibition of ACTH Secretion*. *Endocrine Reviews*, *5*(1), 1–24. https://doi.org/10.1210/edrv-5-1-1

Koolhaas, J. M., Korte, S. M., de Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W., & Blockhuis, H. J. (1999). Coping styles in animals: Current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, *23*, 925–935.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Laskowski, K. L., Chang, C.-C., Sheehy, K., & Aguiñaga, J. (2022). Consistent Individual Behavioral Variation: What Do We Know and Where Are We Going? *Annual Review of Ecology, Evolution, and Systematics*, *53*(1), 161–182. https://doi.org/10.1146/annurev-ecolsys-102220-011451

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P. M., & Meaney, M. J. (1997). Maternal Care, Hippocampal Glucocorticoid Receptors, and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science*, *277*(5332), 1659–1662. https://doi.org/10.1126/science.277.5332.1659

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Maestripieri, D., Higley, J. D., Lindell, S. G., Newman, T. K., McCormack, K. M., & Sanchez, M. M. (2006). Early maternal rejection affects the development of monoaminergic

systems and adult abusive parenting in rhesus macaques (Macaca mulatta). *Behavioral Neuroscience*, *120*(5), 1017–1024. https://doi.org/10.1037/0735-7044.120.5.1017

Malkoc, K., Mentesana, L., Casagrande, S., & Hau, M. (2022). Quantifying Glucocorticoid Plasticity Using Reaction Norm Approaches: There Still is So Much to Discover! *Integrative and Comparative Biology*, *62*(1), 58–70. https://doi.org/10.1093/icb/icab196

Manson, J. H., & Perry, S. (2013). Personality structure, sex differences, and temporal change and stability in wild white-faced capuchins (Cebus capucinus). *Journal of Comparative Psychology*, *127*(3), 299–311. https://doi.org/10.1037/a0031316

McCormack, K., Newman, T. K., Higley, J. D., Maestripieri, D., & Sanchez, M. M. (2009). Serotonin transporter gene variation, infant abuse, and responsiveness to stress in rhesus macaque mothers and infants. *Hormones and Behavior*, *55*(4), 538–547. https://doi.org/10.1016/j.yhbeh.2009.01.009

McEWEN, B. S. (1998). Stress, Adaptation, and Disease: Allostasis and Allostatic Load. *Annals of the New York Academy of Sciences*, *840*(1), 33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x

Meaney, M. J. (2001). Maternal Care, Gene Expression, and the Transmission of Individual Differences in Stress Reactivity Across Generations. *Annual Review of Neuroscience*, *24*(1), 1161–1192. https://doi.org/10.1146/annurev.neuro.24.1.1161

Mentesana, L., & Hau, M. (2022). Glucocorticoids in a warming world: Do they help birds to cope with high environmental temperatures? *Hormones and Behavior*, *142*, 105178. https://doi.org/10.1016/j.yhbeh.2022.105178

Monaghan, P. (2008). Early growth conditions, phenotypic development and environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1497), 1635–1645. https://doi.org/10.1098/rstb.2007.0011

Muniz, L., Perry, S., Manson, J. H., Gilkenson, H., Gros-Louis, J., & Vigilant, L. (2010). Male dominance and reproductive success in wild white-faced capuchins (Cebus capucinus) at Lomas Barbudal, Costa Rica. *American Journal of Primatology*, *72*(12), 1118–1130. https://doi.org/10.1002/ajp.20876

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

Nettle, D., & Bateson, M. (2015). Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151005. https://doi.org/10.1098/rspb.2015.1005

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131343. https://doi.org/10.1098/rspb.2013.1343

Palme, R. (2019). Non-invasive measurement of glucocorticoids: Advances and problems. *Physiology & Behavior*, *199*, 229–243. https://doi.org/10.1016/j.physbeh.2018.11.021

Palme, R., Rettenbacher, S., Touma, C., El-Bahr, S. M., & Möstl, E. (2005). Stress Hormones in Mammals and Birds: Comparative Aspects Regarding Metabolism, Excretion, and Noninvasive Measurement in Fecal Samples. *Annals of the New York Academy of Sciences*, *1040*(1), 162–171. https://doi.org/10.1196/annals.1327.021

Parker, K. J., & Maestripieri, D. (2011). Identifying key features of early stressful experiences that produce stress vulnerability and resilience in primates. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1466–1483. https://doi.org/10.1016/j.neubiorev.2010.09.003

Perry, S. (1996). Intergroup encounters in wild white-faced capuchins (Cebus capucinus). *International Journal of Primatology*, *17*(3).

Perry, S. (1997). Male-Female Social Relationships in Wild White-Faced Capuchins (Cebus capucinus). *Behaviour*, *134*(7/8), 477–510.

Perry, S. (1998). Male-Male Social Relationships in Wild White-Faced Capuchins, Cebus capucinus. *Behaviour*, *135*(2), 139–172.

Perry, S. (2011). Social traditions and social learning in capuchin monkeys (Cebus). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 988–996. https://doi.org/10.1098/rstb.2010.0317

Perry, S. (2012). The Behavior of Wild White-Faced Capuchins: Demography, Life History, Social Relationships, and Communication. In *Advances in the Study of Behavior* (Vol. 44, pp. 135–181). Academic Press. https://doi.org/10.1016/B978-0-12-394288-3.00004-6

Perry, S. E., Barrett, B. J., & Godoy, I. (2017). Older, sociable capuchins (*Cebus capucinus*) invent more social behaviors, but younger monkeys innovate more in other contexts. *Proceedings of the National Academy of Sciences*, *114*(30), 7806–7813. https://doi.org/10.1073/pnas.1620739114

Perry, S., Godoy, I., & Lammers, W. (2012). The Lomas Barbudal Monkey Project: Two Decades of Research on Cebus capucinus. In P. M. Kappeler & D. P. Watts (Eds.), *Long-*

Term Field Studies of Primates (pp. 141–163). Springer. https://doi.org/10.1007/978-3-642-22514-7_7

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P.-O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4051–4063. https://doi.org/10.1098/rstb.2010.0208

Roff, D. (2001). *Life History Evolution* (1st ed.). Sinauer Associates.

Romero, L. M. (2004). Physiological stress in ecology: Lessons from biomedical research. *Trends in Ecology & Evolution*, *19*(5), 249–255. https://doi.org/10.1016/j.tree.2004.03.008

Romero, L. M., & Wikelski, M. (2010). Stress physiology as a predictor of survival in Galapagos marine iguanas. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1697), 3157–3162. https://doi.org/10.1098/rspb.2010.0678

Rosenberg, J., Beymer, P., Anderson, D., Van Lissa, C. j., & Schmidt, J. (2018). tidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. *Journal of Open Source Software*, *3*(30), 978. https://doi.org/10.21105/joss.00978

Royauté, R., & Dochtermann, N. A. (2017). When the mean no longer matters: Developmental diet affects behavioral variation but not population averages in the house cricket (*Acheta domesticus*). *Behavioral Ecology*, *28*(1), 337–345. https://doi.org/10.1093/beheco/arw164

Sanchez, M. M. (2006). The impact of early adverse care on HPA axis development: Nonhuman primate models. *Hormones and Behavior, Translational Topics in Behavioral Neuroendocrinology*, *50*(4), 623–631. https://doi.org/10.1016/j.yhbeh.2006.06.012

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions*. *Endocrine Reviews*, *21*(1), 55–89. https://doi.org/10.1210/edrv.21.1.0389

Schaebs, F. S., Perry, S. E., Cohen, D., Mundry, R., & Deschner, T. (2017). Social and demographic correlates of male androgen levels in wild white-faced capuchin monkeys (Cebus capucinus). *American Journal of Primatology*, *79*(7), e22653. https://doi.org/10.1002/ajp.22653

Schoenemann, K. L., & Bonier, F. (2018). Repeatability of glucocorticoid hormones in vertebrates: A meta-analysis. *PeerJ*, *6*, e4398. https://doi.org/10.7717/peerj.4398

Schoof, V. a. M., & Jack, K. M. (2013). The Association of Intergroup Encounters, Dominance Status, and Fecal Androgen and Glucocorticoid Profiles in Wild Male White-Faced Capuchins (Cebus capucinus). *American Journal of Primatology*, *75*(2), 107–115. https://doi.org/10.1002/ajp.22089

Sheriff, M. J., Dantzer, B., Delehanty, B., Palme, R., & Boonstra, R. (2011). Measuring stress in wildlife: Techniques for quantifying glucocorticoids. *Oecologia*, *166*(4), 869–887. https://doi.org/10.1007/s00442-011-1943-y

Sonnweber, R., Araya-Ajoy, Y. G., Behringer, V., Deschner, T., Tkaczynski, P., Fedurek, P., Preis, A., Samuni, L., Zommers, Z., Gomes, C., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2018). Circadian Rhythms of Urinary Cortisol Levels Vary Between

Individuals in Wild Male Chimpanzees: A Reaction Norm Approach. *Frontiers in Ecology and Evolution*, *6*. https://doi.org/10.3389/fevo.2018.00085

Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259–268. https://doi.org/10.2307/2389364

Stearns, S. C. (2000). Life history evolution: Successes, limitations, and prospects. *Naturwissenschaften*, *87*(11), 476–486. https://doi.org/10.1007/s001140050763

Stedman, J. M., Hallinger, K. K., Winkler, D. W., & Vitousek, M. N. (2017). Heritable variation in circulating glucocorticoids and endocrine flexibility in a free-living songbird. *Journal of Evolutionary Biology*, *30*(9), 1724–1735. https://doi.org/10.1111/jeb.13135

Taff, C. C., Schoenle, L. A., & Vitousek, M. N. (2018). The repeatability of glucocorticoids: A review and meta-analysis. *General and Comparative Endocrinology*, *260*, 136–145. https://doi.org/10.1016/j.ygcen.2018.01.011

Taff, C. C., & Vitousek, M. N. (2016). Endocrine Flexibility: Optimizing Phenotypes in a Dynamic World? *Trends in Ecology & Evolution*, *31*(6), 476–488. https://doi.org/10.1016/j.tree.2016.03.005

Touma, C., & Palme, R. (2005). Measuring Fecal Glucocorticoid Metabolites in Mammals and Birds: The Importance of Validation. *Annals of the New York Academy of Sciences*, *1046*(1), 54–74. https://doi.org/10.1196/annals.1343.006

van Schaik, C. P., & Isler, K. (2012). Life-History Evolution in Primates. In *The Evolution of Primate Societies*. University of Chicago Press.

Weaver, I. C. g., Diorio, J., Seckl, J. R., Szyf, M., & Meaney, M. J. (2004). Early Environmental Regulation of Hippocampal Glucocorticoid Receptor Gene Expression: Characterization

of Intracellular Mediators and Potential Genomic Target Sites. *Annals of the New York Academy of Sciences*, *1024*(1), 182–212. https://doi.org/10.1196/annals.1321.099

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743. https://doi.org/10.1111/brv.12131

Wheeler, B., Tiddi, B., Kalbitzer, U., Visalberghi, E., & Heistermann, M. (2013). Methodological Considerations in the Analysis of Fecal Glucocorticoid Metabolites in Tufted Capuchins (Cebus apella). *International Journal of Primatology*, *34*, 879–898.

Wingfield, J. C. (2013). The comparative biology of environmental stress: Behavioural endocrinology and variation in ability to cope with novel, changing environments. *Animal Behaviour, Special Issue: Behavioural Plasticity and Evolution*, *85*(5), 1127–1133. https://doi.org/10.1016/j.anbehav.2013.02.018

Wingfield, J. C., Maney, D. L., Breuner, C. W., Jacobs, J. D., Lynn, S., Ramenofsky, M., & Richardson, R. D. (1998). Ecological Bases of Hormone—Behavior Interactions: The "Emergency Life History Stage." *American Zoologist*, *38*(1), 191–206. https://doi.org/10.1093/icb/38.1.191

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, *3*(1), 129–137. https://doi.org/10.1111/j.2041-210X.2011.00125.x

Zimmer, C., Taff, C. C., Ardia, D. R., Ryan, T. A., Winkler, D. W., & Vitousek, M. N. (2019). On again, off again: Acute stress response and negative feedback together predict

resilience to experimental challenges. *Functional Ecology*, *33*(4), 619–628. https://doi.org/10.1111/1365-2435.13281

Chapter 3: Estimating glucocorticoid reaction norms from opportunistic fecal samples: a rolling window approach

Abstract

Characterizing individual differences in glucocorticoid secretion from field-collected samples is challenging as samples often represent unknown phases of the physiological response to environmental or social challenges. In fecal glucocorticoid studies, this problem is compounded by excretion delays and uneven sampling across post-stressor intervals. Using 2,427 fecal glucocorticoid samples collected from wild white-faced capuchins (*Cebus imitator*) between 2006 and 2024, I examined whether biologically informative dimensions of hypothalamic-pituitary-adrenal (HPA) axis function could be identified from naturally occurring intergroup encounters.

I describe a generalizable rolling-window approach for identifying where in the post-stressor excretion window individual differences are estimable from opportunistic fecal samples. Applying this approach to white-faced capuchins, I find that a detectable signal occurred 0.5-4.5 hours post-encounter. I then estimated tonic and reactive dimensions of HPA axis activity and found little evidence of covariance between them. The population-average post-encounter effect was positive between approximately 2 and 4.5 hours after intergroup encounters, but the post-encounter effect in the subsequently defined early-response window did not reach conventional significance ($\beta = 0.069$, $p = 0.12$). Guided by this analysis, I estimated two individual-level measures: tonic HPA activity from baseline samples and an early post-encounter responsiveness measure.

Baseline glucocorticoid concentrations showed moderate repeatability (ICC = 0.295), indicating stable among-individual differences in tonic HPA axis activity. Tonic activity and early responsiveness showed little evidence of covariation (r = -0.07), suggesting that they may

capture distinct dimensions of HPA axis function. These results suggest that tonic glucocorticoid activity and acute responsiveness may represent partially independent dimensions of HPA axis function in wild capuchins while highlighting the challenges of estimating endocrine responses to social challenges from opportunistic field data. More broadly, this case study demonstrates the value of empirically characterizing post-stressor sampling windows before extracting individual-level measures of HPA axis activity.

Introduction

Activation of the hypothalamic-pituitary-adrenal (HPA) axis in response to a challenge initiates a coordinated endocrine cascade in which corticotrophin-releasing hormone (CRH) from the hypothalamus stimulates secretion of adrenocorticotrophic hormone (ACTH) from the anterior pituitary, which in turn stimulates glucocorticoid secretion from the adrenal cortex (Romero, 2004; Sapolsky et al., 2000). In white-faced capuchins, as in other primates, cortisol is the primary glucocorticoid of interest. Following an acute challenge, cortisol concentrations rise above pre-challenge levels, reach a peak, and subsequently decline toward tonic levels through regulatory processes including glucocorticoid-mediated negative feedback (Romero, 2004; Sapolsky et al., 2000). This temporal trajectory from tonic activity through activation, peak response, and recovery constitutes the dynamic HPA-axis response to challenge. Importantly, individuals may differ in each component of this response, including tonic activity, the magnitude and timing of activation, and the rate or extent of recovery (Sapolsky et al., 2000).

Individual differences in glucocorticoid secretion are increasingly recognized as meaningful axes of physiological variation in wild animal populations, but characterizing these differences remains both methodologically and conceptually challenging (Cockrem, 2013; Hau et al., 2016; Hau & Goymann, 2015; Malkoc et al., 2022; Taff et al., 2018). Because

glucocorticoids are labile hormones that vary with social, environmental, temporal, and sampling context, single measurements may obscure biologically important differences in endocrine phenotype (Hau & Goymann, 2015; Stedman et al., 2017; Taff & Vitousek, 2016). A reaction norm framework offers a useful lens by modeling glucocorticoid output as a function of context, with individuals differing both in average output across contexts - the intercept or elevation of the reaction norm - and in the magnitude of change across contexts - the slope or plasticity of the reaction norm (Dingemanse et al., 2010; Hau et al., 2022; Malkoc et al., 2022; Sonnweber et al., 2018). Under this framework, stress responsivity is better treated as a context-dependent dimension of stress physiology phenotype, rather than as a fixed trait captured by a single estimate. The glucocorticoid stress response itself can be decomposed into multiple components, including tonic activity, the magnitude or peak of the response following exposure to a stressor, and the subsequent rate or extent of recovery. These components can be difficult to characterize even under controlled conditions and are especially challenging to estimate in free-ranging animals, where repeated sampling across the full response trajectory is rarely possible. Importantly, these components need not necessarily covary - an individual with relatively high tonic glucocorticoid activity may or may not mount a strong acute response to a stressor (Mentesana & Hau, 2022; Romero, 2004; Sapolsky et al., 2000); just as two individuals with similar tonic activity may differ markedly in how they respond to the same challenge.

This distinction maps onto a broader recognition that HPA axis function operates across multiple temporal scales (Keller-Wood & Dallman, 1984; Sapolsky et al., 2000). Tonic or baseline glucocorticoid concentrations integrate current energetic state, reproductive condition, circadian timing, and individual physiology, and may also reveal repeatable among-individual differences when measured repeatedly under comparable conditions (Bonier et al., 2009; Busch

& Hayward, 2009; Romero, 2004; Sapolsky et al., 2000). Acute glucocorticoid responses capture the capacity to mobilize resources in response to immediate challenge (Wingfield et al., 1998). The persistence or recovery of activation following a stressor represents a third condition, reflecting how quickly individuals return toward baseline after perturbation (Romero & Wikelski, 2010; Taff et al., 2018; Zimmer et al., 2019). These components of HPA-axis regulation can vary independently within individuals; consequently, no single glucocorticoid measurement can characterize the dynamic stress response or the broader neuroendocrine phenotype (Malkoc et al., 2022; Taff et al., 2018; Taff & Vitousek, 2016).

Estimating individual reaction norms from field-collected endocrine data presents particular challenges. In captive challenge paradigms, defined stressor timing and serial sample collection (often using blood or plasma samples collected at predetermined intervals) can often be standardized, and the full arc of response and recovery can be traced for each individual. In this way distinct phases of the glucocorticoid response can be characterized including the rising limb, approximate peak, and return toward pre-challenge levels. In long-term field studies, samples are often collected opportunistically and represent uneven coverage of the physiological response curve. Fecal glucocorticoid metabolites (fGCMs) add an additional layer of complexity since circulating cortisol changes appear in feces only after a species-specific excretion lag (Palme, 2019; Palme et al., 2005; Sheriff et al., 2011; Wheeler et al., 2013). Moreover, the timing of fecal glucocorticoid excretion may itself vary among individuals – repeated post-stressor sampling in wild fray mouse lemurs, for example, has revealed substantial individual variation in the timing of peak fecal glucocorticoid metabolite concentrations (Hämäläinen et al., 2014). As a result, samples collected at different times following a stressor may represent different phases of the response (Taff et al., 2018). If individuals are unevenly sampled across

these phases, a single "post-stressor" category conflates response magnitude, timing, and persistence, and estimates of individual "reactivity" may be uninterpretable (Hau et al., 2022; Malkoc et al., 2022; Touma & Palme, 2005; Wheeler et al., 2013). However, when the environmental gradient itself can be quantified continuously (as when precipitation indices serve as a proxy for ecological challenge) a reaction norm approach can sidestep some of these difficulties by modeling glucocorticoid output as a function of that gradient rather than categorizing samples relative to discrete stressor events. Carrera et al. (2025) used this strategy with fecal glucocorticoids from wild white-faced capuchins at Lomas Barbudal, estimating individual reaction norm slopes against a drought risk index and demonstrating that steeper glucocorticoid responses to previous droughts predicted survival across a severe El Niño event.

Here, I apply the reaction-norm framework in an event-centered sense, using time since naturally occurring intergroup encounters to reconstruct individual variation in the temporal dynamics of glucocorticoid activity follow social challenge. Intergroup encounters provide a valuable naturalistic context for studying HPA axis dynamics in wild white-faced capuchins (*Cebus imitator*). Such encounters are socially salient, potentially risky events involving competition over space, food, mates, and group membership (Perry, 1996; Schoof & Jack, 2013), and they are a recurrent feature of capuchin social life (Fragaszy et al., 2004; Perry, 2012), making them ecologically meaningful stressors rather than experimentally imposed disturbances. Because intergroup encounters are not scheduled by the observer, post-encounter fecal samples vary substantially in the time elapsed since the event. This variation creates both a challenge and an opportunity. It complicates simple baseline-versus-post-encounter comparisons but also allows the temporal structure of the post-encounter glucocorticoid response to be examined empirically, provided the sampling distribution is characterized beforehand.

One approach to identifying biologically informative temporal intervals without specifying their boundaries a priori is sliding-window analysis. Sliding-window methods have been developed in ecology to identify critical periods along continuous temporal gradients by systematically shifting the position of a candidate window and evaluating how model support or estimated biological effects change across windows (van de Pol & Cockburn, 2011; van de Pol et al., 2016; Bailey & van de Pol, 2016). Although developed primarily to identify temporal windows of environmental exposure, this framework is applicable more generally when biological effects are expected to vary along a continuous time axis.

In this study, I adapt a sliding-window approach previously used to identify temporally localized biological signals in ecological and endocrine data (Vitousek et al., 2022; van de Pol & Cockburn, 2011; van de Pol et al., 2016; Bailey & van de Pol, 2016) to estimate individual differences in HPA axis function from opportunistic fecal samples, using data from wild white-faced capuchins (*Cebus imitator*) at Lomas Barbudal, Costa Rica as a case study. Rather than using sliding windows to identify the historical period over which an environmental exposure predicts a biological outcome, I shift a fixed-width window across time since a discrete stressor (intergroup encounters) to identify where a post-stressor fGCM signal is detectable and where among-individual variation in response magnitude can be estimated. While the specific window boundaries reported are particular to *Cebus imitator* and intergroup encounters, the rolling-window characterization, stacking of baseline and post-event samples, and diagnostic assessment of random slope estimability is potentially transferable to other field systems.

Methods

Fecal glucocorticoid sample collection, processing, and assay

Fecal samples were collected opportunistically immediately following defecation and stored in a cooled thermos until returned to the field laboratory, a lag time of approximately 1–12 hours. Samples were then frozen and stored at −20°C. On a monthly basis, samples were allowed to thaw and come to room temperature before being mechanically homogenized and dried in a conventional oven at 80–115°C for 2–3 hours. Prior validation in this population confirms that glucocorticoid concentrations derived from wet and heat-dried fecal samples are comparable using this method (Beehner et al., 2022). Once dried, undigested material (insect parts, bone, fur, pith, and seeds) was removed, and the remaining material was ground to a fine powder and stored in WhirlPak bags at ambient temperature until shipment to the University of Michigan, where samples were stored at −20°C until extraction.

Fecal glucocorticoids were extracted from fecal powder using an 80% ethanol solution (prepared with deionized water). The standard extraction used 0.10 g of fecal powder in 2 ml of ethanol solution, however when sample volume was insufficient, a subset of samples was extracted using 0.05 g of fecal powder in 1 ml of solution, preserving the powder-to-solvent ratio. The validity of this reduced-weight protocol was confirmed in a parallel extraction experiment in which a subset of samples was processed at both weight endpoints and concentrations compared; results are reported in the Supplementary Materials (Section S5). After adding the ethanol solution, samples were vortexed for 10 minutes prior to centrifugation for an additional 10 minutes at 3000-4000 rpms. The supernatant was then transferred to a labeled microcentrifuge tube and stored at −20°C until assay.

Glucocorticoids were assayed using the DetectX® Cortisol Enzyme Immunoassay Kit (Arbor Assays, K003-H), which has been validated for use in this population (Beehner et al., 2022). This kit uses a competitive ELISA format in which cortisol in the sample competes with a

cortisol-peroxidase conjugate for binding to a monoclonal antibody; signal is read at 450 nm. The assay sensitivity is 27.6 pg/mL, with a standard curve range of 50–3,200 pg/mL. Cross-reactivity with corticosterone and cortisone is low (1.2% each). Although solvolysis is often recommended prior to fecal extract assay to improve antibody binding, Beehner et al. (2022) demonstrated high agreement between solvolyzed and non-solvolyzed extracts in white-faced capuchins ($r = 0.98$), and this step was therefore omitted. Samples were diluted in assay buffer across a range of 1:2.5 to 1:200 to bring concentrations within the standard curve. All samples were run in duplicate; samples for which the coefficient of variation between duplicate wells exceeded 15% were re-run. Concentrations were calculated in ng/g dry fecal weight using MyAssays software, accounting for sample dilution. A fecal pool from a non-experimental animal was extracted and serially diluted to generate high (1:10) and low (1:100) controls. The low control produced values that fell outside the reliable range of the standard curve on multiple plates (straddling the asymptote), so intra- and inter-assay CVs are reported only for the high control. Intra-assay CV (within a plate) averaged 9.97% ($n = 60$ duplicate pairs). Inter-assay CV across Z plates averaged 20%. Prior to all analyses, fecal glucocorticoid concentrations were log-transformed.

During data processing, one batch of samples (January 21, 2024) exhibited anomalously low glucocorticoid concentrations. Subsequent analyses (see Supplementary Materials, Section S4) indicated that this pattern was consistent with a processing artifact related to drying conditions. These samples ($n = 8$) were excluded from final analyses.

Baseline vs. post-IGE sample classification

Each fecal glucocorticoid sample was classified as either a post-intergroup encounter (post-IGE) sample or a tonic sample based on the timing of confirmed intergroup encounters relative to sample collection. In brief, intergroup encounters occur when two social groups with

overlapping home ranges meet. These are often tense encounters that carry the risk of physical aggression between members of opposing groups. Such aggression may lead to severe wounding. These events typically resolve by one (or both) groups fleeing the area of encounter. Encounter identification and verification procedures are described in depth in the Supplementary Materials (Section S1). Matching was performed within focal group - each sample was compared only to confirmed encounters recorded for the same social group, using the collection date and time of the sample and the verified start time of each encounter.

A sample was classified as post-IGE if a confirmed intergroup encounter occurred within the 24 hours preceding collection. Where multiple confirmed encounters fell within this window, the sample was matched to the most recent encounter. The elapsed time between encounter start and sample collection was retained as a covariate in subsequent analyses of post-IGE samples. A sample was classified as tonic if no confirmed intergroup encounter could be identified within the 24 hours preceding collection for that individual's group. I use the term *tonic* here rather than *baseline* glucocorticoid activity because field-collected fGCM measures integrate HPA-axis activity over time and cannot establish a standardized, unstimulated physiological baseline. Here tonic activity refer specifically to glucocorticoid output measured outside the defined post-IGE response window.

Covariates

Reproductive state

fGCM concentrations are known to vary across reproductive state within this population (Carrera et al., 2025); therefore reproductive state was included as a covariate in all models. Reproductive state was determined for each fecal sample *post hoc* using demographic records maintained in the Lomas Barbudal long-term database. Briefly, females were classified as

pregnant based on an estimated conception date calculated as 158 days prior to parturition, consistent with the average gestation length reported for white-faced capuchins (Carnegie et al., 2011). Pregnancy was further divided into early (first half) and late (second half) stages based on the estimated parturition date. Females were classified as nursing for the 12-month period following the birth of a surviving infant; if an infant died within its first year of life, nursing status was applied only for the interval between birth and infant death. Females were classified as cycling if they had completed the nursing period or had previously weaned an infant but had not yet conceived again, and as prereproductive if they were younger than four years of age and had no history of pregnancy or nursing. A postreproductive classification was applied to females who had been continuously cycling for more than approximately 2.25 years without conceiving. Because fGC concentrations in this population have been found to differ across pregnancy stages but not between cycling, nursing, and prereproductive females (Carrera et al., 2025), these latter categories were grouped as a single "not pregnant" reference level in all statistical models. 'Male' was included as its own separate level. Reproductive state was included as a covariate in all models to account for its known influence on glucocorticoid output.

Alpha status

Alpha male status was coded as a binary covariate for use in fecal glucocorticoid models, following the operationalization described in Chapter 2. Briefly, a male was coded as alpha (1) during periods of confirmed, settled alpha tenure and as non-alpha (0) at all other times, including periods of contested or uncertain dominance. Alpha status was included as a fixed effect covariate in HPA axis models to account for its known association with elevated tonic glucocorticoid concentrations in male capuchins (Jack et al., 2014; Schaebs et al., 2017). Because alpha status is included as a covariate in HPA axis models, the individual-level HPA

axis parameters extracted for downstream analyses reflect dispositional HPA axis variation beyond what is attributable to the elevation in GCs associated with holding alpha rank.

Time of day

Sample collection time in minutes past midnight was included as a continuous covariate to account for circadian variation in cortisol secretion. Fecal glucocorticoid concentrations in this population are higher in samples collected earlier in the day, reflecting the lag between diurnal peaks in circulating cortisol and their appearance in feces (Carrera et al., 2025).

Current precipitation

Rolling 30-day total precipitation derived from ERA5 daily reanalysis data was included as a continuous covariate to account for the influence of current ecological conditions on cortisol output. Glucocorticoid concentrations in this population are elevated during drier periods, reflecting the energetic demands of reduced food availability (Carrera et al., 2025). This covariate captures short-term environmental state at the time of sample collection and is distinct from the multi-year rainfall deviation measures used as ELA predictors.

Statistical Analysis

All analyses were conducted in R (version 4.3.1) using the lme4 package (Bates et al., 2015) for linear mixed-effects model fitting, with denominator degrees of freedom and p-values computed via Satterthwaite's method as implemented in lmerTest (Kuznetsova et al., 2017). Intraclass correlation coefficients (ICCs) were computed using the performance package (Lüdecke et al., 2021). All continuous predictors were scaled to a mean of zero and standard deviation of one prior to model fitting. Models were fit using restricted maximum likelihood (REML) with the bobyqa optimizer.

Tonic HPA axis activity

Tonic HPA axis activity was estimated from fecal glucocorticoid samples for which no confirmed intergroup encounter occurred within the preceding 24 hours. This approach

quantifies stable, non-challenge-related individual differences in cortisol output. Log-transformed fGC concentration was modeled as a function of time of day (minutes past midnight, scaled), 30-day cumulative precipitation (scaled), alpha status (binary), and reproductive state (categorical: male [reference], nonpregnant, nursing, pregnant), with a random intercept for individual identity. The intraclass correlation coefficient (ICC) was computed from the fitted model to quantify the proportion of residual variance attributable to stable individual differences. Individual-level best linear unbiased predictors (BLUPs) for the random intercept were extracted and retained as the measure of tonic HPA axis activity.

Characterizing the post-IGE excretion window

Given that intergroup encounters at Lomas Barbudal cluster around 08:00 and field days end at approximately 18:00–19:00, fecal samples collected more than approximately 12 hours post-IGE are collected the following morning, after a full sleep-wake cortisol nadir and awakening response. These next-day samples cannot be meaningfully interpreted as reflecting the acute post-encounter cortisol trajectory. All post-IGE analyses were therefore restricted to same-calendar-day samples collected between 15 minutes and 12 hours post-encounter.

To empirically characterize when, if ever, a post-IGE fGC signal is detectable under natural field sampling conditions (i.e., post-IGE samples differ significantly from tonic samples), I applied a rolling-window analysis to the same-day post-IGE dataset. At each of 33 window centers ranging from 2 to 10 hours post-IGE (advancing in 0.25-hour steps), I constructed a dataset comprising all tonic samples and the subset of same-day post-IGE samples falling within +/-2 hours of the window center (4-hour total bandwidth). A linear mixed-effects model was fit to each windowed dataset with log-transformed fGC as the outcome, a binary context indicator (0 = tonic, 1 = post-IGE) as a fixed effect, time of day, age, 30-day precipitation, alpha status,

and reproductive state as covariates, and individual identity as a random intercept and random slope on the context indicator. The fixed-effect estimate for the context term (β_{ctx}) indexes the population-average post-IGE elevation at each window center, and the random slope variance quantifies between-individual variation in the magnitude of that elevation. Models that converged to a singular solution were refit with random intercepts only. This analysis is exploratory; adjacent windows heavily overlap, and no single window's nominal significance should be interpreted as an independent hypothesis test. Accordingly, window selection was based on the temporal pattern of effect estimates, biological plausibility, sampling coverage, and estimability of among-individual variance rather than on selecting the window producing the smallest nominal *p*-value. This distinction is important because exploratory sliding-window procedures can otherwise inflate support for apparently optimal temporal windows through repeated model fitting (van de Pol et al., 2016; Bailey & van de Pol, 2016).

Response window selection

The rolling-window results were used to define two biologically informed post-IGE sampling windows for subsequent phenotyping models. An *early/rising window* (0.5–4.5 hours post-IGE) was selected to capture the ascending limb of the fGCM excretion curve, corresponding to the interval where the population-average context effect was positive and appreciable between-individual variation in the context effect could be estimated. Among these candidate windows, the 2.5-hour centered window had the largest population-average context effect and was therefore selected for subsequent characterization. The late-response window (8-12 hours post-IGE) was also examined. Due to sparse sampling within this window (median = 1 post-IGE sample per individual), I was unable to estimate reliable individual differences in late-

window glucocorticoid deviations. Full results from this exploratory analysis are provided in the Supplementary Materials.

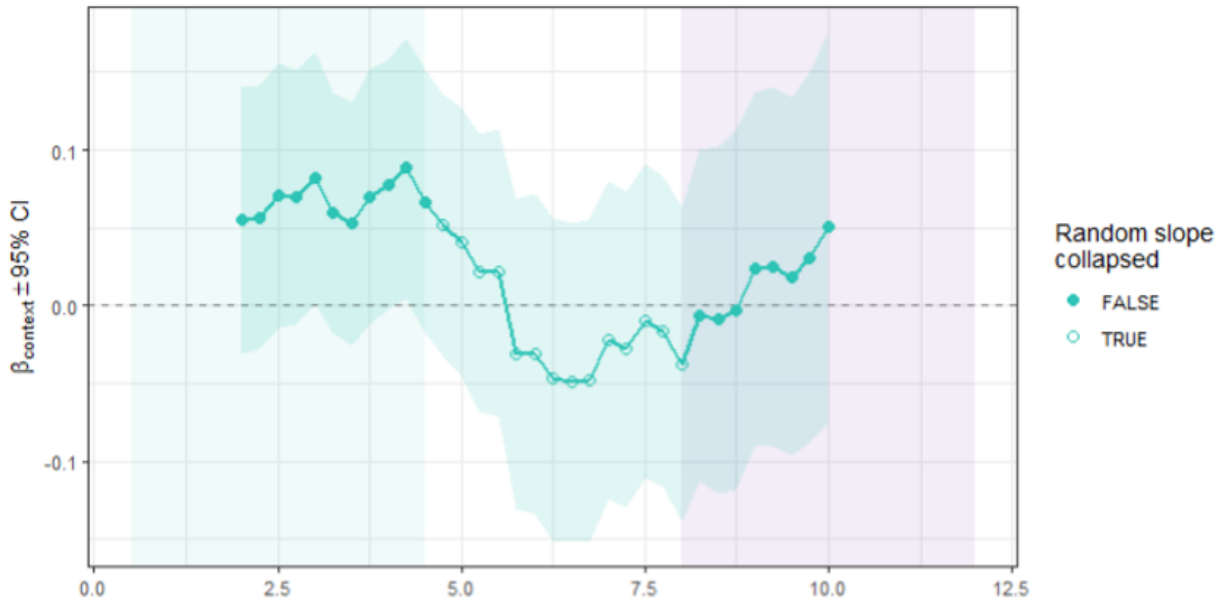


Figure 3.1. Rolling-window analysis of post-IGE fGC signal. Plot shows the population-average post-IGE context effect. Blue shading = early/rising window (0.5-4.5h); purple shading = late-response window (8-12h). Open circles indicate windows where the random-slope model was singular and was refit with random intercepts only.

Early/rising post-IGE responsiveness.

Individual differences in early post-IGE responsiveness were estimated from a stacked dataset comprising all tonic samples and the subset of same-day post-IGE samples falling within the early/rising window (0.5–4.5 hours post-encounter). Log-transformed fGC was modeled with a binary context indicator as a fixed effect, alongside hours post-IGE centered at the window midpoint, time of day, age, 30-day precipitation, alpha status, and reproductive state as covariates. The random-effects structure included a random intercept and random slope on the context indicator for individual identity. The fixed effect of context tests the population-average early-window shift in fGC, and the random slope variance estimates whether individuals differ reliably in the magnitude of their early post-IGE response. Individual-level BLUPs for the context slope were extracted as a measure of early/rising HPA axis responsivity. Because context

was binary, these estimates quantify individual differences in the magnitude of the tonic-to-post-IGE shift; they do not estimate individual rates of glucocorticoid change over elapsed time following an encounter.

Late-window response (exploratory)

An analogous stacked model was fit using tonic samples and same-day post-IGE samples from the late-response window (8–12 hours post-encounter), with the same fixed- and random-effects structure as the early/rising model (hours post-IGE centered at the late-window midpoint). However, due to sparse sampling within this window (median = 1 post-IGE sample per individual), I was unable to estimate reliable individual differences in late-window glucocorticoid deviations. Full results and interpretation are provided in the Supplementary Materials (S3).

Relationships among HPA axis dimensions

Pairwise Pearson correlations among the three extracted components of HPA axis activity (tonic, early/rising, and late-response) were computed to assess their independence. Because these measures were extracted as BLUPs, correlations among them should be interpreted as diagnostics of whether the extracted individual-level estimates contain overlapping information, rather than as definitive estimates of the latent covariance structure of HPA axis function. A multivariate mixed model estimating covariance among tonic and reactive components directly would be preferable in principle, but the sparse post-IGE sampling within biologically appropriate windows limited the feasibility and stability of such models. A strong correlation between the tonic component and a reactive component would indicate that the two measures capture overlapping individual-level variance and cannot be interpreted as independent

dimensions of HPA axis function. This diagnostic step determines which components can defensibly be carried forward into downstream analyses.

Results

Individual differences in HPA axis function

The dataset used in this analysis comprised 2,084 tonic fGCM samples from 130 individuals (median = 13 samples per individual, range 1–55) collected over a mean within-individual span of 3.14 years (median = 0.86, range = 0 - 17.22 years) and 343 same-day post-IGE samples from 95 individuals (median = 3 samples per individual, range 1–13), collected over a mean within-individual span of 2.33 years (median = 1.13, range = 0 – 14.71 years) and between 15 minutes and 12.1 hours post-encounter.

Tonic HPA axis activity

The tonic-only model (N = 2,055 observations from 124 individuals after listwise deletion of missing covariates) revealed stable between-individual differences in fGCM concentrations. The random intercept variance was 0.093 (SD = 0.305), yielding an adjusted intraclass correlation coefficient of 0.295, indicating that approximately 30% of the residual variance in tonic fGC is attributable to consistent individual-level differences. Among the fixed effects, time of day ($\beta = -0.047$, $p < 0.001$), 30-day cumulative precipitation ($\beta = -0.121$, $p < 0.001$), alpha status ($\beta = 0.216$, $p = 0.002$), and reproductive state (all levels $p < 0.001$ relative to the male reference category) were significant predictors. Pregnancy was associated with the largest elevation in fGC relative to males ($\beta = 1.093$), followed by nonpregnant females ($\beta = 0.631$) and nursing females ($\beta = 0.592$). Individual-level BLUPs for the random intercept were extracted and retained as the measure of tonic HPA axis activity for downstream analyses.

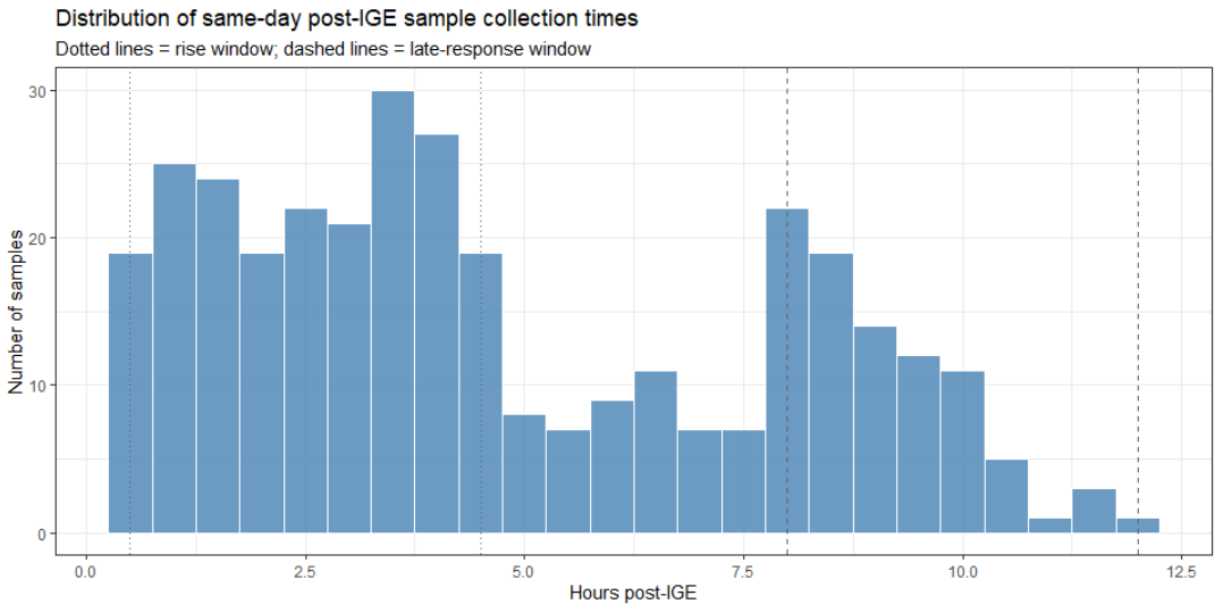


Figure 3.2. Distribution of same-day post-IGE sample collection times.

Rolling-window characterization of the post-IGE signal

The rolling-window analysis revealed a time-varying post-IGE context effect across the same-day sampling range (Figure 3.1). The fixed-effect context estimate (β*ctx*) was positive across the 2–4.5-hour collection window, with its largest estimate occurring at a window center of approximately 3-4.5 hours, with a maximum at 4.25 hours (β = 0.089, SE = 0.043). Beyond this point, the confidence intervals widened substantially, and the data did not support a consistent positive elevation thereafter (Table S3.1). Fecal samples collected 3–4.25 hours post-IGE likely reflect cortisol secreted approximately 0–1.25 hours after the encounter corresponding to the early ascending limb of the acute cortisol response. These specific temporal boundaries reflect the combination of the 3-4 hour gut transit time previously estimated for this species (Damm & Gault, 2020; Wheeler et al., 2013) and intergroup encounters at this site cluster around 08:00, such that same-day sampling extends approximately 10-11 hours before field operations conclude. Because adjacent rolling windows overlap substantially, estimates from successive windows are not independent hypothesis tests. Rather, the rolling-window analysis

was used to characterize how the magnitude of the context effect and among-individual variation in responsivity changed across the post-IGE sampling period and to inform selection of the response window used in subsequent analyses.

Between window centers of 4.75 and 8.0 hours, random slope models converged to singular solutions and were refit with random intercepts only, indicating that the data could not support estimation of between-individual variation in the context effect within these windows. In the non-singular windows at the ascending (2-2.5 hours) and descending (8.25-10 hours) tails, random slope variance was appreciable (ascending: 0.028-0.032; descending: 0.018-0.036), whereas near the descriptive population peak (3-4.5 hours), random slope variance was substantially smaller (0.002-0.006).

Early/rising post-IGE responsiveness.

The early/rising stacked model comprised 2,244 observations: 2,054 tonic and 190 post-IGE samples from 71 individuals collected within the 0.5–4.5-hour window (median = 2 post-IGE samples per individual, range 1–7; 52 individuals contributed ≥2 samples). The population-average context effect indicated a trend toward elevated fGC in the post-IGE window relative to tonic levels ($\beta = 0.069$, SE = 0.043, $p = 0.12$). Although this effect did not reach conventional significance in the selected early-response model, the rolling analysis indicated that the magnitude of the population effect varied with window position, reaching a nominally significant maximum in the later-positioned window centered at 4.25 h. The selected early window was chosen to support estimation of individual differences in responsiveness rather than to maximize population-level statistical significance. The within-window time variable (hours centered at the window midpoint) was not significant ($\beta = 0.033$, $p = 0.289$). The random slope variance for the context term was 0.033 (SD = 0.181), and the model-estimated correlation between the random

intercept and context slope was −0.05, indicating that early responsiveness is essentially uncorrelated with tonic HPA axis activity within this model. Although the population-average early-window context effect was positive, it did not reach conventional significance. The early/rising measure should therefore be interpreted not as evidence for a uniform population-wide post-IGE elevation, but as an individual-difference estimate derived from the window where the rolling analysis suggested the strongest and most biologically plausible post-encounter signal. Individual BLUPs for the context slope were extracted as the estimate of early/rising HPA axis responsiveness.

Late-window analysis (exploratory)

The late-window stacked model (8-12 hours post-encounter) showed no population-average context effect and, critically, was estimated from very sparse per-individual sampling (median = 1 post-IGE sample). Therefore, I do not extract individual-level BLUPs for this window. Full results are reported in the Supplementary Materials (Section S3).

Table 3.1. Fixed effects from early/rising and late-response stacked models. Rise model: N = 2244 from 124 individuals (190 post-IGE from 71 individuals).

Fixed effect	**Estimate**	**SE**	**T**	***p***
Intercept	2.1023	0.0528	39.813	0.000
Context (post-IGE)	0.0689	0.0436	1.579	0.1195
Hours post-IGE	0.0327	0.0308	1.061	0.2890
Time of day	-0.0483	0.0103	-4.686	0.000
30-day precipitation	-0.1193	0.0106	-11.285	0.000
Age	-0.0059	0.0035	-1.699	0.0901
Alpha status	0.2485	0.0678	3.664	0.000

Nonpregnant female	0.6420	0.0727	8.824	0.000
Pregnant female	1.0949	0.0773	14.165	0.000

Table 3.2. Variance components from early/rising and late-response stacked models. Rise model: N = 2244 from 124 individuals (190 post-IGE from 71 individuals).

Component	Variance	SD/Corr
Individual (intercept)	0.0987	0.3141
Individual (slope)	0.0326	0.1805
Intercept-slope correlation	NA	-0.0488
Residual	0.2158	0.4645

Relationships among HPA axis components

Pairwise Pearson correlations among the three extracted HPA axis components were computed to assess their independence. Because individuals contributing only a single post-IGE sample to a given window have context-slope BLUPs that are heavily shrunk toward the population mean and carry minimal individual-level information, correlations involving reactive components were computed after restricting to individuals with ≥ 2 post-IGE samples in the relevant window. Full-sample correlations are also reported here for transparency.

Table 3.3. Pairwise Pearson correlations among tonic and rising responsiveness components. Filtered correlations restrict to individuals with ≥ 2 post-IGE samples in the relevant window to reduce shrinkage bias in context-slope BLUPs.

Sample	N	R	95% CI lower	95% CI upper	*p*
Full sample	123	-0.121	-0.292	0.058	0.183
≥ 2 post-IGE	52	-0.073	-0.339	0.204	0.608

Tonic HPA axis activity and rising responsiveness showed little evidence of covariance in the filtered sample (r = -0.073, p = *0.*608, N =52) and in the full sample (r = -0.121, p = *0.*183, N = 123). However, as discussed below, the precision of the reactive component estimate limits the strength of inference that can be drawn from this null correlation.

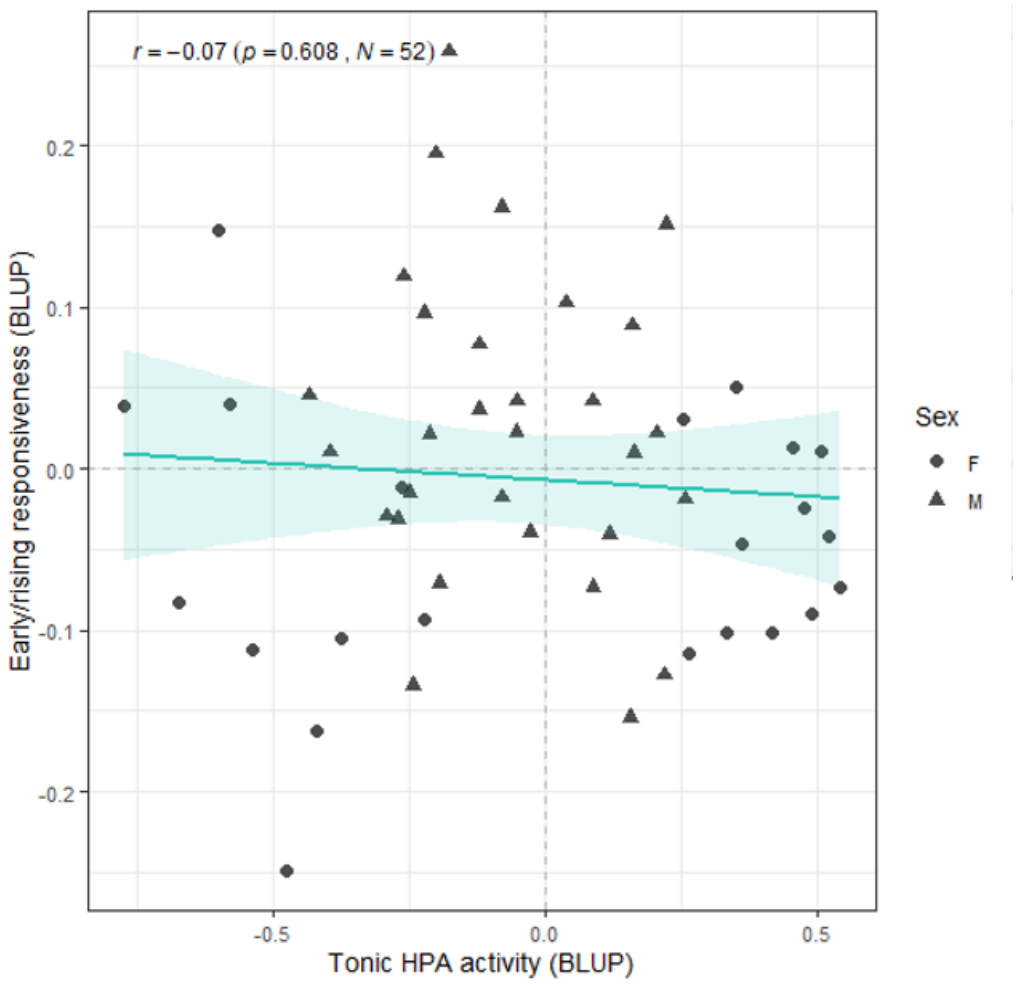


Figure 3.3. Independence of tonic and reactive HPA axis components. Tonic HPA axis activity BLUPs (random intercept from the tonic-only model) are plotted against early/rising responsiveness BLUPs (random slope on the context indicator from the 0.5–4.5-hour post-IGE stacked model) for individuals contributing ≥2 post-IGE samples in the rise window (N = 52). Points are shaped by sex (circles = female, triangles = male). The fitted regression line with 95% confidence band shows no association between the two components (r = −0.07, p = 0.608), showing little evidence of association between the two components.

Summary and measures carried forward.

Tonic HPA axis activity, estimated from the tonic-only model with an adjusted ICC of 0.295, is the primary individual-difference measure carried forward into subsequent chapters. The early/rising HPA axis responsiveness component provides an independent estimate of individual differences in acute post-encounter cortisol deviation and is retained as a secondary measure, with the caveat that it is estimated from a limited number of post-IGE samples per individual (median = 2) and that the population-average context effect did not reach conventional significance ($p = 0.12$). Null associations involving the early/rising responsiveness measure should be interpreted as inconclusive rather than as evidence against a relationship, given the modest per-individual sample sizes available in the biologically appropriate excretion window.

Discussion

Tonic activity and acute responsiveness show little evidence of covariation.

This study makes two contributions: one methodological, one empirical. Methodologically, I demonstrate a rolling-window procedure for empirically characterizing when within the post-stressor excretion window fGCM concentrations are elevated relative to tonic levels and when individual differences can be reliably estimated. This procedure addresses the fact that post-stressor samples collected at different times may represent different phases of the physiological response, a well-recognized challenge in field endocrinology (Hau et al., 2022; Malkoc et al., 2022; Taff et al., 2018), by providing a data-driven way to identify estimable dimensions rather than relying on post-hoc assumptions about gut transit time. Empirically, using white-faced capuchins as a case study, I find that tonic HPA axis activity and early post-encounter responsiveness represent separable dimensions of individual glucocorticoid variation. The correlation between the tonic component (estimated from tonic samples) and the early/rising

responsiveness component (estimated from the ascending limb of the post-IGE fGCM curve) was indistinguishable from zero ($r = -0.07$, $N = 52$ individuals with ≥2 rise-window samples), and this independence was consistent across sample-size thresholds ($r = -0.002$ at ≥3 samples). Knowing an individual's tonic cortisol level provides essentially no information about how that individual's cortisol deviates from tonic levels within the first 4.5 hours following an intergroup encounter, and vice versa.

This decoupling is consistent with evidence from other vertebrates that tonic and stress-induced glucocorticoid activity can show different patterns of among-individual variation and need not covary within populations (Taff et al., 2018; Mentesana & Hau, 2022). Experimental studies of captive capuchins provide further support, demonstrating distinct regulation of circulating cortisol across circadian, ACTH-stimulated, and stress-induced contexts. Studies of *Cebus* and *Sapajus* have documented circadian variation in plasma cortisol and ACTH responsiveness, cortisol responses to repeated capture and venipuncture, and multihour plasma cortisol dynamics following repeated blood-sampling stressors (Dettmer et al., 1996; Lahoz et al., 2007; Torres-Farhan et al., 2008; Wheeler et al., 2013).

A critical caveat to this interpretation concerns the precision with which the reactive phenotype was estimated. The early responsiveness BLUPs were derived from a median of only two post-IGE samples per individual (range: 1-7), and simulation analyses parameterized from the observed data indicated that with two post-IGE samples per individual, the correlation between true and estimated individual slopes was approximately 0.46 (Section S6). Under these conditions, even a true correlation between tonic and reactive dimensions would be substantially attenuated toward zero (Houslay & Wilson, 2017). The near-zero correlation observed here is therefor consistent both with genuine independence and with the presence of a true relationship

that is masked by measurement error in the reactive dimension. Distinguishing these possibilities will require studies with denser post-stressor sampling and more precise estimation of individual response trajectories. One feasible approach may be to conduct repeated all-day focal follows of individual monkeys, collecting every fecal sample produced throughout the observation period. This design could provide substantially denser sampling across the subsequent excretion window and permit more precise estimation of individual response trajectories. Such targeted sampling would be particularly valuable if designed around the sampling requirements identified by the power analysis presented in the Supplementary Information. I therefore interpret the near-zero correlation as preliminary evidence that tonic and reactive measures may capture partially distinct aspects of HPA axis function, rather than as definitive demonstration of their independence.

The present results provide empirical support for the distinction between dissociable components of HPA axis activity in a wild primate. This finding parallels work in Carrera et al. (2025), who used a reaction norm framework to show that individual differences in glucocorticoid slopes across a drought gradient predicted survival during an El Niño event independently of intercept differences - a result that would be undetectable if intercept and slope were treated as a single trait.

The practical consequence for this dissertation is that early-life adversity, behavior, and life-history outcomes could relate to tonic activity and responsiveness in different ways, or in opposite directions, without contradiction. A developmental exposure that elevates tonic cortisol need not simultaneously alter acute reactivity, and an individual that is physiologically "reactive" to social stressors may or may not also carry high tonic levels of cortisol. Treating these as separate components in subsequent chapters is therefore an empirically motivated decision.

Individual differences concentrate at response boundaries, not peak magnitude.

An unexpected finding from the rolling-window analysis was that the windows supporting estimation of between-individual variation in the post-IGE context effect were the ascending and descending tails, not the window containing the population-average peak. At the peak (3-4.5 hours post-encounter), random slope variance was an order of magnitude smaller (0.002-0.006) than at the tails (0.018-0.036). This pattern suggests that white-faced capuchins are relatively homogeneous in the magnitude of their peak fGCM response to intergroup encounters but differ substantially in the timing of response onset and the rate of recovery.

This finding, if replicated with serial sampling designs, would have implications for how field studies conceptualize and measure individual differences in HPA axis function. Most studies that estimate "stress reactivity" from fecal samples rely on a single post-stressor measurement, often collected at a fixed interval presumed to capture the peak (e.g., 3-4 hours post-event). The present results suggest that such an approach may capture the dimension of HPA axis function that shows the least individual variation (peak magnitude), while missing the dimensions that show the most variation (onset timing and recovery rate), an interpretation that may help explain why repeatability estimates for glucocorticoid reactivity can vary widely across studies and why some studies find low or non-significant repeatability of stress-induced glucocorticoid concentrations (Baugh et al., 2014; Schoenemann & Bonier, 2018; Taff et al., 2018). Single-point peak sampling may therefore systematically underestimate the true extent of individual differences in stress physiology and may misidentify which individuals are most physiologically distinct.

However, an alternative explanation is equally plausible: the signal-to-noise ratio may be highest near the population peak because all individuals are elevated, and the model may lack the

power to partition variance when the contextual effect is large relative to individual differences. Under this view, the larger random slope variance at the tails reflects greater relative influence of noise rather than greater biological variation in response timing. Distinguishing these alternatives will require serial sampling designs that can trace individual response trajectories.

Individual differences in recovery dynamics could arise in part from variation in glucocorticoid negative-feedback regulation, which contributes to termination of HPA-axis activation. This pattern is consistent with a regulatory architecture in which negative feedback efficiency is a primary axis of individual variation, while maximal secretory capacity is relatively conserved (Sapolsky et al., 2000; Zimmer et al., 2019). However, the present data cannot identify the physiological mechanisms underlying variation in response trajectories. Differences in initial activation, adrenal responsiveness, feedback regulation, and fecal excretion dynamics could all contribute to the observed among-individual variation.

Characterizing components of stress-response phenotypes from opportunistic field data

Applying the rolling-window procedure to the data revealed several patterns relevant to field studies more broadly. First, the population-average post-IGE context effect varied across the post-encounter window: positive in the 2–4.5 hour range ($\beta \sim 0.07$–0.09), near-zero from 5–8 hours, and inconsistently signed beyond 8 hours (although the magnitude of the population-average context effect is modest, the consistency of positive estimates across adjacent rolling windows and their alignment with the expected 3-4 hour gut transit time in this species suggest that this is a genuine biological signal rather than assay noise). Second, the ability to estimate between-individual variance in the context effect tracked this temporal structure. Models were singular (unable to support a random slope) across the 4.75–8-hour range, but non-singular with appreciable slope variance at the ascending and descending tails. Third, approximately one-third

of nominally "post-IGE" samples (those falling within 12–24 hours of an encounter) were collected the following morning, after a full sleep-wake cycle and cortisol awakening response. These next-day samples cannot be meaningfully interpreted as reflecting acute HPA axis reactivity, yet they would be included in any analysis using a fixed 24-hour post-stressor window.

The practical implication is that field studies should empirically characterize the excretion window before extracting measures of responsiveness, rather than relying on post-hoc assumptions about transit time. The specific boundaries identified are likely specific to *Cebus imitator* and intergroup encounters, but the general approach of stacking tonic and post-stressor samples, sliding a window across the post-stressor interval, and evaluating where the context effect and its random-slope variance are estimable is directly transferable to other species and stressor types.

Limitations

Several features of this analysis constrain the strength of inference that can be drawn from the components of HPA axis activity estimated here. The most robustly estimated measure was tonic HPA axis activity, which was based on 2,055 tonic observations from 124 individuals and showed moderate repeatability (ICC = 0.295). Even so, this measurement accounted for less than one-third of the residual variance in tonic fGC. The remaining variance likely reflects a combination of within-individual day-to-day fluctuation, measurement error intrinsic to fGCM assays, and unmeasured contextual factors, such as social interactions, foraging success, and weather conditions, which influence cortisol on a given day but are not stable properties of the individual. An ICC of 0.295 is sufficient to support extraction of individual-level estimates, but it

also means that the tonic BLUP for any given individual is estimated with meaningful uncertainty.

This uncertainty is greater for the early/rising responsiveness component. Although the rolling-window analysis identified the early post-encounter interval as the most biologically plausible window for estimating acute responsiveness, the population-average context effect in the rising window did not reach conventional significance ($p = 0.12$). In addition, individual estimates were based on a limited number of post-IGE samples per individual: the median was two samples, with 52 of 71 individuals contributing at least two samples. The random slope variance was non-zero (0.033), and the independence of the rising component from tonic activity was well supported, but the reliability of any individual's rise BLUP remains modest. When these BLUPs are carried forward as predictors in subsequent analyses, the measurement error they contain is expected to attenuate coefficient estimates toward zero (Houslay & Wilson, 2017). Null results involving the rising component should therefore be interpreted as potentially reflecting limited precision in the phenotype estimate, rather than as strong evidence that acute responsiveness is unrelated to the outcome in question.

To quantify these constraints, I conducted a simulation study parameterized from the observed data (N = 70 individuals, 13 tonic samples per individual, random slope variance = 0.033, context $\beta = 0.069$; full details in Section S6). Varying the number of post-IGE samples per individual from 1 to 20 across 500 simulated datasets, I found that models failed to estimate a random slope (singularity) in 22% of simulations at 1 sample per individual and 9% at 2 samples, dropping below 5% only at 3 or more samples. The correlation between true and estimated individual slopes was 0.34 at 1 post-IGE sample (the late-window median), 0.46 at 2 samples (the rise-window median), and did not exceed 0.70 until approximately 10 samples per

individual. Power to detect the population-average context effect at the observed magnitude (β = 0.069) was only 32% at 2 samples per individual and plateaued around 60% even with 20 samples, indicating that more individuals, not only more samples per individual, would be needed to reliably detect effects of this magnitude.

A related limitation is that all measures of HPA axis activity here were extracted as BLUPs, which are point estimates that do not propagate uncertainty around each individual's random effect. A more fully principled approach would carry this uncertainty into downstream models, for instance by jointly estimating the HPA axis phenotype model and the outcome model within a single Bayesian framework, or by treating HPA axis function as a latent variable in a structural equation model. With the current sample sizes, particularly for the reactive component, such approaches were not feasible. The BLUP-based strategy adopted here is common in animal behavior and behavioral ecology studies and is partially justified by the moderate ICC for tonic HPA axis activity, but the resulting conservative bias in downstream effect estimates should be kept in mind when evaluating associations reported in the following chapter.

The current dataset was especially limited for evaluating post-encounter recovery dynamics. Sampling in the 8–12-hour post-encounter window was sparse, with a median of one post-IGE sample per individual. As a result, I was unable to estimate reliable individual differences in late-window glucocorticoid deviations or recovery rate. Full results from this exploratory analysis are provided in the Supplementary Materials (Section S3), but the late-window measure should not be treated as a primary measure of HPA axis activity.

The inference that individuals may differ more in onset timing or recovery dynamics than in peak magnitude is also a population-level statistical inference, not a direct observation of individual response curves. The finding that random slope variance was larger at the tails than

near the apparent peak derives from mixed-model variance estimates across the population. The interpretation that individuals may differ in onset timing and recovery rate while converging on similar peak magnitudes should therefore be treated as a hypothesis generated by the rolling-window analysis, one that requires confirmation with serial sampling designs.

Finally, the measured components estimated here characterize individual differences in HPA axis function across the full span of each individual's adult sampling history. The tonic BLUP is therefore a long-term average that assumes rank-order stability across years. To the extent that individuals' relative cortisol levels shift over the lifespan, for example with aging, changes in dominance status, or reproductive history, this averaging may obscure developmentally dynamic patterns. Including alpha status and reproductive state as covariates mitigates this concern to some degree but does not eliminate it.

Conclusion

This study establishes that HPA axis function in wild white-faced capuchins is not a single trait but a set of partially independent physiological dimensions. Tonic cortisol activity and acute post-encounter responsiveness showed little evidence of covariance, suggesting that an individual's tonic endocrine state may provide little information about how that individual's HPA axis responds to social challenge. This pattern is consistent with broader comparative evidence that tonic and stress-induced glucocorticoid measures can capture partially independent dimensions of HPA-axis regulation (Taff et al., 2018). However, this null result should be interpreted cautiously given the modest reliability of the reactive phenotype estimate. More broadly, the rolling-window approach described here provides a generalizable, data-driven method for identifying where in the post-stressor excretion window individual differences are estimable from opportunistic fecal samples. Rather than assuming that any post-stressor sample

collected within a fixed interval (e.g., 24 hours) is equally informative, researchers can use this procedure to empirically characterize the temporal structure of the glucocorticoid response in their own study system. These methodological tools, together with the empirical demonstration that tonic and reactive components can be independently estimated from fecal samples, provide a foundation for future studies examining the causes and consequences of individual variation in HPA axis function.

Supplementary Information: Estimating glucocorticoid reaction norms from opportunistic fecal samples: a rolling window approach.

S1. Identification of Intergroup Encounters

Intergroup encounters were identified by searching all available observational data for text entries containing any of the following terms or their common misspellings: “intergroup,” “inter-group”, “intgrp”, “intergrp”, “intergpoup”, “intrgroup”, and “IG encounter”. Because a single intergroup encounter may generate multiple observational entries (e.g. “intergroup begins”, descriptions of participants, “intergroup ends”), observations were clustered into candidate encounters. Observations from the same focal group on the same calendar day occurring within 60 minutes of one another were assigned to the same encounter cluster. The earliest timestamp within each cluster was designated as the encounter start time.

This search yielded 6,317 individual observations grouped into 3,427 candidate encounter clusters. Each observation was then classified into one of four categories based on manual review of the free-text content. The four classification categories and their decision rules are described below.

S1.1 Classification Categories and Decision Rules

YES (confirmed intergroup encounter; N = 5,156 classified observations). An observation was classified as YES if the text described an intergroup encounter that was occurring at or near the time of the observation, or that had clearly just occurred and was still affecting group behavior. Diagnostic features included: direct statements of an ongoing encounter (e.g. “C INTERGROUP”, “intergroup with RR”, “intergrouping”); descriptions of behavioral responses consistent with an active encounter (running, fleeing, alarm calling, threats toward another group); use of the progressive tense (“intergrouping”); data collection disruptions caused by the encounter (“abort due to intergroup”, “lost monos during intergroup”, “dropped focal for intergroup”); references to dictaphone or audio recordings of the encounter; specific

identification of the opposing group (“intergroup with Pelon”, “intergroup with RR”); and immediate aftermath observations clearly belonging to the same event cluster (e.g. “males returning from intergroup”, “intergroup over”, “everyone tense after intergroup” when occurring within the same time cluster).

NO (not an intergroup encounter, N = 419 classified observations). An observation was classified as NO if it referenced the word “intergroup” (or a variant) but did not describe an encounter occurring at or near the time of observation. Five primary contexts motivated NO classifications: (1) Retrospective references to a past encounter (“intergroup last night”, “after the intergroup yesterday”); (2) Wound or injury checks conducted after a prior encounter (“wound from intergroup”, “scars from intergroup with WY”, “blood on him from the intergroup”); (3) Explicit negation (“no intergroup”, “not an intergroup”, “doesn’t seem like an intergroup”); (4) Future or hypothetical references (“intergroup tomorrow”, “waiting for intergroup”, “hope no intergroup”; (5) Meta-commentary (reflections on field conditions or counting past encounters).

AMBIGUOUS (uncertain; N = 566 classified observations). An observation was classified as AMBIGUOUS when the text was consistent with an intergroup encounter but lacked sufficient detail to confirm one occurred, or that could link the timestamp associated with the text entry with an encounter. Common patterns included: isolated question-mark references with no supporting behavioral description (“intergroup?”); hedged language without confirming details (“possibly intergroup”, “there could be an intergroup”); inferred encounters not directly witnessed (“there must have been an intergroup”, “may have been an intergroup this morning”); observations noting individuals missing after an encounter whose occurrence was itself uncertain; and cases where the observer explicitly expressed uncertainty about whether the

behavioral response reflected an actual intergroup encounter or some other event ("thought it was an intergroup but think maybe not"). Aftermath observations that confirmed an encounter had occurred but could not contribute temporal precision to encounter start time estimation were also classified as AMBIGUOUS (e.g., "still haven't seen X since intergroup," "monos at same spot since intergroup this morning").

GEOGRAPHIC_REF (geographic reference; N = 176 manually classified observations). An observation was classified as GEOGRAPHIC_REF when the text referenced "Intergroup Island," a named landmark within the study site. These observations consisted of location records ("CL INTERGROUP ISLAND"), distance and bearing measurements from the landmark ("30M N INTERGROUP ISLAND"), or descriptions of animals resting or foraging at the location ("monos resting at ST tree at Intergroup Island"). The term "camino intergroup" (a trail name) was also classified in this category.

S1.2 Encounter-Level Aggregation

Observation-level classifications were aggregated to the encounter level using a hierarchical rule: an encounter cluster was classified as YES if any constituent observation was classified as YES; otherwise, AMBIGUOUS if any observation was AMBIGUOUS; otherwise as NO; and if none of the above as GEOGRAPHIC_REF. This approach reflects the fact that a single encounter cluster often contains a mixture of active encounter descriptions, retrospective notes about the same event, and uncertainty expressions. The presence of at least one confirmed observation was treated as sufficient evidence that the encounter occurred.

This procedure yielded 2,751 confirmed intergroup encounters, 381 ambiguous candidate encounters, 197 non-encounters, and 108 geographic references across the full observation period.

To supplement and validate this observational-derived dataset, I cross-referenced encounter records against an independent set of GPS-logged intergroup encounters recorded by field assistants using handheld GPS devices between 2009 and 2021. When an intergroup encounter was observed in the field, assistants marked the event on the GPS device, producing a timestamped and geolocated record of the encounter. For records collected prior to 2013, the GPS devices did not reliably capture time-of-day, so only the date component of these records was used for cross-referencing.

Cross-referencing the GPS dataset against the observational encounter summary yielded two corrections. First, 29 encounters that had been classified as ambiguous in the observational dataset were corroborated by a GPS record from the same group on the same date, these were reclassified as confirmed. For these upgraded encounters, the encounter start time from the original observational record was retained, as all 29 fell within the pre-2013 period when GPS timestamps were unreliable. Second, 22 GPS-recorded encounters occurring after 2013, and therefore carrying reliable timestamps, had no corresponding entry in the observational encounter summary. These were added to the dataset as confirmed encounters with the GPS-recorded timestamp serving as the encounter start time. The final encounter dataset used for computing the time-post-intergroup-encounter covariate and confirming classification as 'tonic' samples in all subsequent hormone analyses comprised a total of 2,802 confirmed encounters spanning 2002–2024.

S2. Rolling-window analysis post-IGE context effect

Table S3.1. Rolling-window analysis of post-IGE context effect. For each window center (in hours post-encounter), the table shows the number of post-IGE samples entering the model (n_post), the number of individuals contributing both tonic and post-IGE samples (n_both), fixed-effect estimate for the context indicator (post-IGE vs. tonic), standard error, p-value, variance associated with the random slope on context (reactive_var) and variance associated with the random intercept (tonic_var) from the linear mixed model fit to tonic samples and post-IGE

samples within ±2 hours of the window center. This table is provided for transparency; the analysis is exploratory and descriptive, and no correction for multiple comparisons is applied.

center	n_post	n_both	beta	se	p	reactive_var	tonic_var	Singular
2	187	77	0.0490	0.0411	0.237	0.0214	0.0985	FALSE
2.25	187	73	0.0572	0.0433	0.192	0.0322	0.0985	FALSE
2.5	190	70	0.0717	0.0434	0.104	0.0322	0.0986	FALSE
2.75	189	67	0.0701	0.0417	0.0983	0.0205	0.0985	FALSE
3	184	69	0.0819	0.0412	0.0519	0.0167	0.0986	FALSE
3.25	172	67	0.0598	0.0394	0.135	0.00320	0.0985	FALSE
3.5	166	67	0.0528	0.0397	0.189	0.00228	0.0982	FALSE
3.75	153	68	0.0701	0.0418	0.0986	0.00539	0.0979	FALSE
4	157	70	0.0784	0.0408	0.0594	0.00341	0.0976	FALSE
4.25	143	69	0.0886	0.0425	0.0411	0.00370	0.0972	FALSE
4.5	139	68	0.0671	0.0430	0.123	0.00306	0.0971	FALSE
4.75	134	65	0.0521	0.0430	0.226	NA	0.0999	TRUE
5	129	66	0.0410	0.0437	0.348	NA	0.0984	TRUE
5.25	119	65	0.0218	0.0452	0.630	NA	0.0988	TRUE
5.5	111	62	0.0220	0.0469	0.638	NA	0.0990	TRUE
5.75	95	54	-0.0305	0.0507	0.547	NA	0.0990	TRUE
6	88	50	-0.0302	0.0526	0.566	NA	0.0986	TRUE
6.25	90	47	-0.0464	0.0525	0.377	NA	0.0977	TRUE
6.5	93	46	-0.0485	0.0519	0.351	NA	0.0964	TRUE
6.75	91	47	-0.0467	0.0525	0.374	NA	0.0961	TRUE
7	94	46	-0.0202	0.0518	0.697	NA	0.0969	TRUE
7.25	96	46	-0.0265	0.0516	0.607	NA	0.0961	TRUE
7.5	98	48	-0.00803	0.0512	0.875	NA	0.0958	TRUE
7.75	102	50	-0.0150	0.0505	0.767	NA	0.0957	TRUE
8	100	50	-0.0360	0.0511	0.481	NA	0.0950	TRUE
8.25	103	51	-0.00319	0.0539	0.953	0.0179	0.0975	FALSE
8.5	99	51	-0.00636	0.0567	0.911	0.0290	0.0970	FALSE
8.75	97	49	-0.000208	0.0585	0.997	0.0356	0.0971	FALSE
9	93	48	0.0264	0.0576	0.648	0.0233	0.0973	FALSE
9.25	91	47	0.0280	0.0585	0.634	0.0249	0.0972	FALSE
9.5	90	46	0.0215	0.0583	0.714	0.0216	0.0972	FALSE
9.75	87	45	0.0331	0.0601	0.585	0.0265	0.0973	FALSE
10	82	44	0.0532	0.0637	0.408	0.0368	0.0972	FALSE

S3. Late-window analysis of post-encounter fGCMs

S3.1 Rationale and limitations

The rolling-window analysis described in the main text suggested that the population-average post-IGE context effect declined toward zero beyond approximately 5.5 hours post-encounter, with models failing to support random slope estimation across much of the 4.75-8-hour range. However, to fully characterize the post-encounter excretion window and to assess

whether any late-window signal might be detectable with sufficient sampling, I fit an exploratory stacked model using tonic samples and same-day post-IGE samples from the late-response window (8-12 hours post-encounter).

Crucially, the late-window analysis presented here is severely limited by sparse per-individual sampling. The median number of late-window post-IGE samples per individual was 1 (range 1-4), and only 21 of 46 individuals contributed ≥2 samples. Under these conditions, random slope estimates are heavily shrunk toward the population mean and carry minimal individual-level information. The results presented below are therefore not interpretable as reliable measures of individual-differences and are not carried forward into any downstream analyses. They are presented here for transparency and to illustrate the sample size requirements for estimating post-recovery glucocorticoid dynamics from opportunistic field data.

S3.2 Methods

The late-window stacked model comprised all tonic samples (N = 2,084) and same-day post-IGE samples collected within the 8–12-hour post-encounter window (N = 82 samples from 46 individuals). Log-transformed fGC concentration was modeled with a binary context indicator (0 = tonic, 1 = post-IGE) as a fixed effect, alongside hours post-IGE centered at the late-window midpoint (10 hours), time of day, age, 30-day precipitation, alpha status, and reproductive state as covariates. The random-effects structure included a random intercept and random slope on the context indicator for individual identity. Models were fit using restricted maximum likelihood (REML) with the bobyqa optimizer in the lme4 package (Bates et al., 2015).

S3.3 Results

The late-window stacked model converged without singularity. The population-average context effect was null (β = 0.035, SE = 0.082, p = 0.673; Table S3.1), indicating no detectable elevation of fGC in the late post-IGE window relative to tonic levels. The within-window time

variable (hours since encounter, centered at 10 hours) also showed no detectable effect (β = -0.017, SE = 0.059, $p = 0.77$).

The random slope variance for the context term was 0.039 (SD = 0.197; Table S3.2), which was larger than the random slope variance in the early/rising model (0.033). However, the model-estimated correlation between the random intercept and the context slope was strongly negative (−0.42), indicating that the model could not cleanly distinguish tonic levels from late-window deviations. This is expected when post-IGE samples are sparse relative to tonic samples (tonic observations constituted approximately 96% of the late-window stacked dataset).

Table S3.2. Fixed effects from the late-window stacked model (N = 2,166 observations from 125 individuals; 82 post-IGE samples from 46 individuals).

Predictor	β	SE	df	t	p
Intercept	2.847	0.058	118.2	49.08	<0.001
Context (post-IGE vs. tonic)	0.035	0.082	54.5	0.43	0.673
Hours post-IGE (centered)	-0.017	0.059	78.3	-0.29	0.770
Time of day (scaled)	-0.051	0.010	1793.0	-5.00	<0.001
Age (scaled)	0.012	0.036	109.1	0.33	0.744
30-day precipitation (scaled)	-0.128	0.009	1886.2	-13.59	<0.001
Alpha status (yes)	0.203	0.084	77.2	2.42	0.018
Reproductive state (vs. male)					
└ Nonpregnant female	0.624	0.081	97.8	7.74	<0.001
└ Nursing female	0.582	0.107	108.7	5.42	<0.001
└ Pregnant female	1.085	0.154	151.1	7.05	<0.001

Table S3.3. Variance components from the late-window stacked model.

Random Effect	Variance	SD	Correlation
Individual (intercept)	0.115	0.339	
Individual (context slope)	0.039	0.197	
Intercept-slope correlation			-0.42
Residual σ^2	0.186	0.431	
ICC (intercept)	0.338		
ICC (context slope)	0.115		

S3.4 Assessment of individual-difference reliability

To evaluate whether the late-window model could support extraction of individual-level BLUPs despite the sparse sampling, I examined how the correlation between late-window BLUPs and tonic HPA axis activity (the best-estimated component) varied as a function of sample inclusion threshold. If the late-window model were reliably estimating individual differences, I would expect the correlation to be relatively stable across thresholds.

Figure S3.1 shows that the correlation between the late-window context slope BLUP and the tonic HPA axis component changed depending on the inclusion criterion. The loss of statistical significance at ≥3 samples and the attenuation of the correlation coefficient from -0.85 to -0.24 indicates that the full-sample correlation was driven primarily by shrinkage in individuals with only a single late-window sample. When the analysis is restricted to individuals with sufficient data to anchor their own slope estimates (≥3 samples), the correlation is no longer

distinguishable from zero.

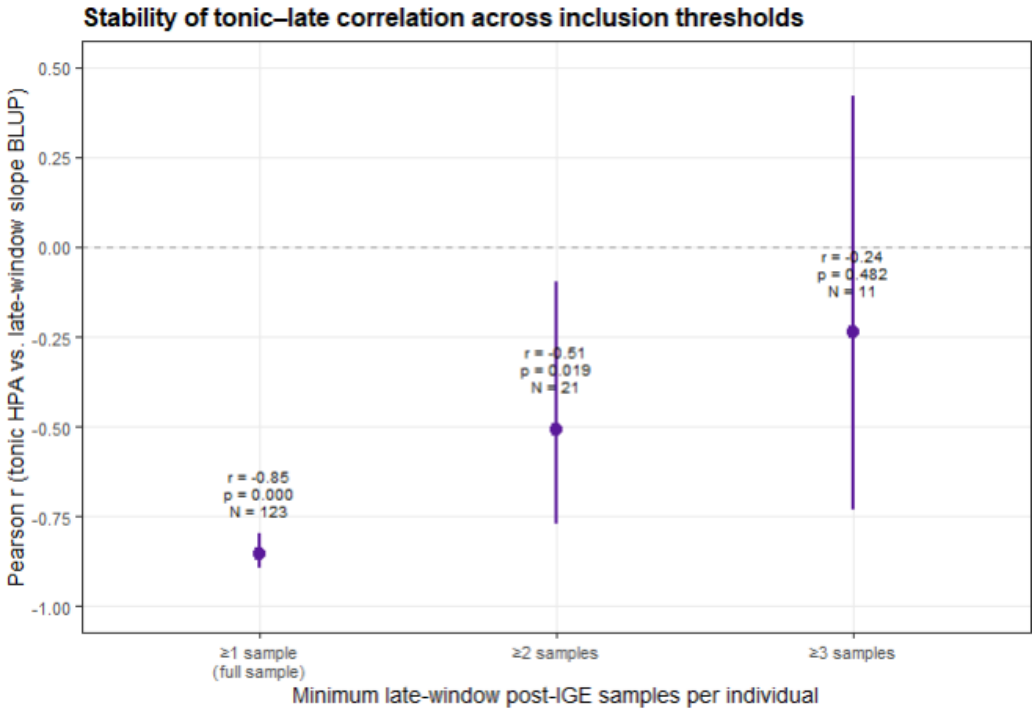


Figure S3.1. Stability of the tonic-late correlation as a function of sample inclusion threshold. Points show the Pearson correlation coefficient (with 95% confidence intervals) between tonic HPA axis activity BLUPs and late-window context slope BLUPs, calculated for subsets of individuals meeting different minimum numbers of late-window post-IGE samples. The correlation attenuates and becomes non-significant at ≥3 samples, indicating that the full-sample correlation is an artifact of shrinkage in sparsely sampled individuals.

S3.5 Interpretation

The late-window analysis demonstrates that there is no population-average elevation of fGC in the 8–12-hour post-encounter window relative to tonic levels. This is consistent with the rolling-window analysis in the main text, which showed the context effect declining toward zero beyond approximately 5.5 hours. It suggests that the fGCM trace of an acute cortisol pulse following an intergroup encounter is largely resolved within 8 hours, at least at the population level.

Furthermore, the late-window model could not reliably estimate individual differences in post-encounter glucocorticoid deviations. The combination of sparse per-individual sampling (median = 1 post-IGE sample) and the moderately strong negative intercept-slope correlation (-

0.42) indicates that the model cannot cleanly separate tonic level differences from late-window deviations. Under these conditions, any extracted BLUPs would be dominated by shrinkage to the population mean rather than reflecting genuine individual variation.

Finally, if the late window captures the return toward tonic levels, individuals who have fully recovered should show late-window fGC that converges on their personal tonic level, yielding a context slope near zero regardless of tonic height. The negative correlation instead indicates that high-tonic individuals tend to show negative deviations from their tonic levels in the late window, while low-tonic individuals tend to show positive deviations. This pattern would be consistent with individual differences in negative feedback efficiency: high-tonic individuals engaging more rapid or complete HPA axis suppression following an acute response, potentially overshooting tonic levels before re-equilibrating, while low-tonic individuals show slower recovery or more sustained post-encounter activation.

S3.6 Implications for field endocrinology studies

The late-window analysis serves as a cautionary example for field studies attempting to estimate individual differences in post-stressor glucocorticoid recovery from opportunistic fecal samples. Estimating a random slope (i.e., individual variation in the effect of a stressor) requires sufficient within-individual replication across both tonic levels and post-stressor conditions. While simulation studies have shown that reliable random slope estimation is possible with as few as 3-5 observations per individual under ideal conditions (e.g., Houslay & Wilson, 2017; Martin et al., 2011), the present data for the late-window analysis falls far below this threshold. For researchers designing field studies with the goal of estimating individual differences in post-stressor glucocorticoid responses from fecal samples, these results suggest that:

1. The post-stressor excretion window must be empirically characterized for the species and stressor of interest.

2. Serial sampling within that window (multiple samples per individual per stressor event) may be necessary to estimate recovery dynamics.

3. Opportunistic sampling protocols that yield a median of 1-2 post-stressor samples per individual are unlikely to support reliable estimation of individual differences in recovery dynamics.

These challenges are not insurmountable, but they require study designs that prioritize repeated post-stressor sampling over broader cross-sectional coverage.

S4. Identification of a potential drying-temperature artifact

During routine laboratory use in late January 2024, the oven temperature dial was observed to be above the standard drying temperature, although the timing and duration of this change was not documented. The standard drying condition for fecal samples in this assay workflow was 95°C. Based on the observed dial setting, a higher temperature of 122°C was identified as a plausible alternative condition that may have been applied inadvertently. Elevated drying temperature could plausibly reduce measured GC concentrations through thermal degradation, volatilization, or altered extraction efficiency (Gholib et al., 2018). I therefore conducted a controlled experiment to directly compare these two drying conditions and examined mean batch-level residuals for anomalies that might align with the inadvertent change in drying temperature.

S4.1 Sample selection and paired design

A subset of fecal samples was selected for validation using a paired design. For each sample, two aliquots were prepared and processed in parallel. For each pair, one aliquot was dried at 95°C and the other at 122°C. Following drying, all samples were processed using the same GC extraction and assay procedures applied throughout the study. Assay outputs consisted

of replicate concentration measurements, and the mean concentration across replicates was used for analysis. Paired aliquots were compared using assay runs at matched dilutions.

For each paired sample, I calculated mean GC concentrations under both temperature conditions and quantified the effect of elevated temperature as both a ratio and percent change.

Table S3.4: **Paired-sample comparison.**

SID	DILUTION	95C	122C	122/95	% CHANGE
45299	15x	4340.00	613.50	0.141	-85.9%
45300	15x	463.45	234.15	0.505	-49.5%
45330	5x	658.70	236.33	0.359	-64.1%
45330	15x	330.88	152.99	0.462	-53.8%
45333	15x	922.35	153.57	0.166	-83.4%
453331	5x	136.01	136.70	1.005	+0.5%
45334	5x	237.70	171.55	0.722	-27.8%

S4.2 Experimental outcome

6 of 7 paired comparisons showed lower measured GC concentrations under the 122C drying condition (see Table S3.4). The mean reduction was 52.0%, with a median reduction of 53.8%, indicating a strong and consistent downward bias under elevated drying temperature.

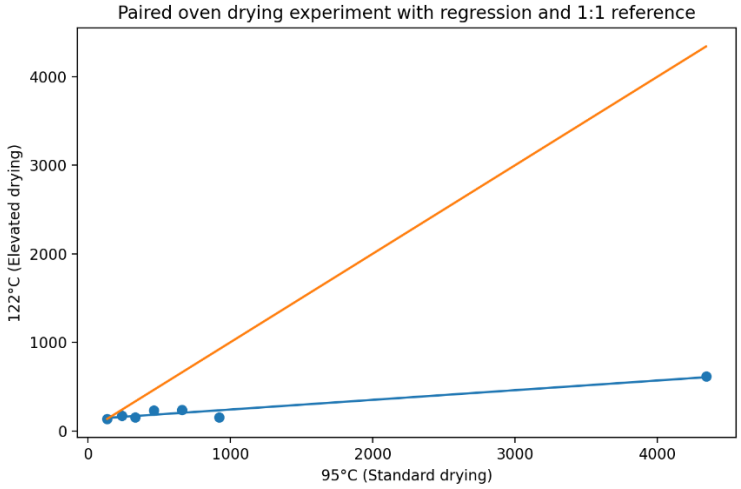


Figure S3.2. Effect of elevated drying temperature on fGCM concentrations. Paired aliquots from the same fecal samples were dried at 95°C (standard) or 122°C (elevated) and processed using identical extraction and assay procedures. Each point represents a matched sample pair. The solid line shows the fitted regression, and the diagonal line represents the 1:1 relationship (no effect of temperature). All points fall below the 1:1 line, indicating systematically lower measured concentrations under elevated drying temperature.

S4.3 Statistical identification of batch-level anomalies

To distinguish potential processing artifacts from expected biological variation, I used a linear mixed-effects modeling approach to explicitly partition variation in GC concentrations into stable between-individual difference, seasonal variation, and residual batch-level deviations. GC concentrations were log-transformed prior to analysis to improve normality. Log GC concentrations were included as the response variable in the model, with a spline function on date added as a fixed effect and individual identity as a random effect. Processing date was converted to a continuous variable (days since first processing date) and modeled using a natural cubic spline with four degrees of freedom, allowing for nonlinear but smooth temporal trends while avoiding over-fitting of shorter-term fluctuations.

After fitting the model, I extracted residuals as reflections of deviations from the expected GC concentrations after accounting for both seasonal trends (captured by processing date) and individual identity. To identify batch-level anomalies, residuals were aggregated by processing date. For each batch, I calculated the mean residual, standard error of the mean, and the 95% CI. Batches with mean residuals substantially below zero represent processing dates where measured GC concentrations were systematically lower than expected, consistent with a potential drying temperature artifact.

S4.4 Interpretation

Because the spline term captures smooth temporal variation, abrupt deviations in batch-level residuals are unlikely to reflect changes in GC concentrations that accompany seasonal change. Instead, these deviations may be interpreted as candidate processing artifacts. The batch processed on 21 January 2024 exhibited a large negative mean residual (~-0.7 log units), corresponding to an approximately 50% reduction relative to expectations. A similarly strong deviation was observed on 22 February 2024, corresponding to the controlled high-temperature

validation experiment, providing an internal positive control for the expected magnitude of a drying-temperature artifact.

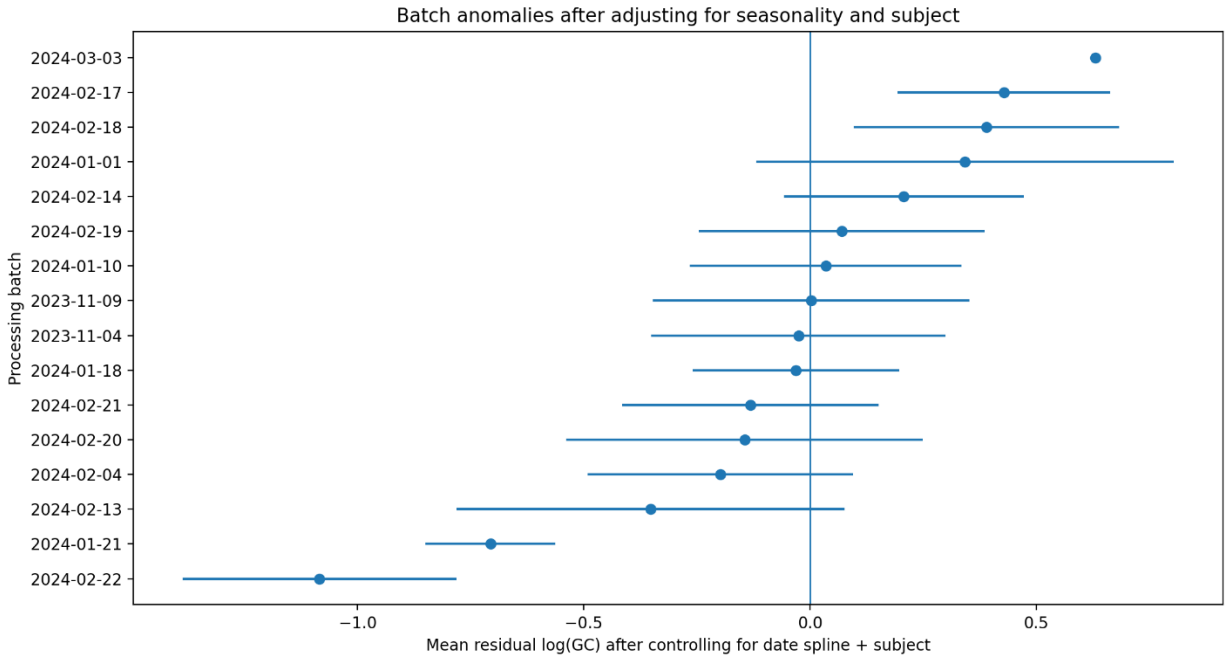


Figure S3.3. Batch-level deviations in fGCM concentrations after controlling for seasonality and subject identity. Mean residual log-transformed GC concentrations are shown for each processing batch following a linear mixed-effects model including a spline term for processing date and a random intercept for subject identity. Points represent batch means and horizontal lines indicate 95% confidence intervals. The vertical line at zero represents the expected value after accounting for temporal and individual-level variation. Most batches cluster around zero, consistent with expected seasonal trends, whereas the 21 January 2024 batch shows a pronounced negative deviation (~50% lower than expected). The 22 February 2024 batch, corresponding to the controlled high-temperature validation experiment, exhibits an even stronger negative deviation. These results indicate that the January 21 batch reflects a discrete processing anomaly consistent with the effects of elevated drying temperature.

S5. Validation of low-weight fecal extracts

The standard fecal extraction protocol used in this study targets a dried fecal weight of 0.15g, with a working minimum of 0.10g. However, a substantial fraction of samples (particularly those collected following intergroup encounters) fell below this 0.10g threshold. To determine whether these low-weight samples could be retained without introducing systematic bias in measured GC concentrations, I conducted a paired validation experiment comparing standard-weight and low-weight extracts of the same samples.

S5.1 Sample selection and paired design

Eight fecal samples from two adult males collected between 2007 and 2012 were selected for validation. For each sample, two aliquots were prepared: a standard-weight aliquot ("A") targeting the protocol weight of 0.15g (mean 0.168g, range 0.158-0.181g) and a low-weight aliquot ("B") well below the protocol minimum (mean 0.053g, range 0.048-0.059g). Standard-weight aliquots were extracted in 2ml of 80% ethanol solution, while low-weight aliquots were extracted in 1ml of 80% ethanol solution. Both aliquots from each pair were assayed using identical procedures, at the same plate dilution (15x), and on the same plate where possible. Paired aliquots were compared by computing within-pair ratios (B/A) and percent change, and by conducting a paired t-test on log-transformed ng/g values.

S5.2 Experimental outcome

Across the eight pairs, low-weight aliquots produced GC concentrations that were statistically indistinguishable from their standard-weight counterparts (Table S5.1). The mean ratio of low-weight to standard-weight concentrations was 0.944 (95% CI [0.816, 1.090]), and a paired t-test on log ng/g values revealed no significant difference between conditions ($t(7) = -0.95$, $p = 0.374$). Six of the eight pairs showed lower measured concentrations under the low-weight condition and two showed higher, with within pair percent changes ranging from -30.2% to +24.5%. The mean percent change across pairs was -4.4% (median -3.5%), indicating no consistent directional bias.

SID	WEIGHT A (G)	WEIGHT B (G)	A NG/G	B NG/G	RATIO B/A	% CHANGE
500	0.181	0.051	84.27	82.46	0.978	−2.2%
739	0.178	0.059	33.35	27.24	0.817	−18.3%
1695	0.171	0.048	74.85	79.67	1.064	+6.4%
2092	0.168	0.056	58.10	57.76	0.994	−0.6%

6638	0.160	0.048	243.19	218.42	0.898	−10.2%
9422	0.165	0.052	52.27	36.48	0.698	−30.2%
9638	0.158	0.057	128.32	122.16	0.952	−4.8%
9647	0.163	0.053	93.94	116.92	1.245	+24.5%

Table S3.5. Paired-sample comparison of standard-weight (A) and low-weight (B) fecal extracts.

S5.3 Interpretation

These results indicate that fecal samples extracted at masses well below the 0.10g protocol minimum yielded GC concentrations comparable to standard-weight extracts, with no detectable systematic bias. The sample size (n = 8 pairs) is modest, and the 95% confidence interval on the mean ratio (0.82–1.09) does not rule out small systematic effects in either direction. However, the absence of a consistent directional shift, combined with the fact that the largest within-pair deviations were of comparable magnitude in both directions, supports the inclusion of low-weight samples in subsequent analyses. Consequently, 114 samples with a fecal weight of 0.05g – 0.1g were extracted in 1ml of 80% ethanol solution and retained in the HPA axis modeling analysis within this chapter.

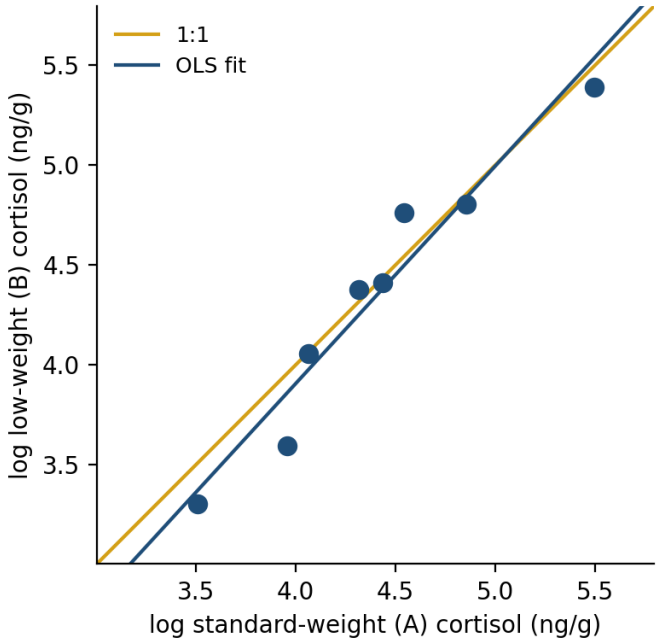


Figure S5.1 Validation of low-weight fecal extracts. Eight fecal samples were each split into a standard-weight aliquot (~0.16 g, "A") and a low-weight aliquot (~0.05 g, "B") and assayed at matched plate dilutions (15×). Each point represents a matched sample pair on the log scale. The gold line indicates perfect agreement (1:1), and the blue line shows the fitted OLS regression (slope = 1.09, intercept = −0.45, r = 0.97). Points fall closely along the 1:1 line, indicating that low-weight aliquots reproduce the cortisol concentrations measured from standard-weight aliquots of the same sample.

S6. Simulation: sample size requirements for random slope estimation

S6.1 Rationale

The analyses presented in this chapter highlight the practical challenge of estimating individual differences in post-stressor glucocorticoid responses (using random slopes) from opportunistic fecal samples. Proper estimation requires sufficient within-individual replication across both tonic and post-stressor conditions. In the present dataset, the median number of post-IGE samples per individual was 2 in the early/rising window and 1 in the late window. To provide concrete guidance on how many post-stressor samples per individual are needed to reliably estimate random slopes under conditions like those encountered in this study, I conducted a simulation study parameterized from the observed data.

S6.2 Methods

The data-generating process was parameterized from the early/rising stacked model. I simulated datasets comprising 70 individuals, each contributing 13 tonic samples (the observed median) and a varying number of post-IGE samples. Individual-level random effects (intercepts and context slopes) were drawn from a bivariate normal distribution with variance–covariance parameters matching the observed estimates: random intercept variance = 0.098, random slope variance = 0.033, intercept–slope correlation = −0.05, and residual variance = 0.186. The population-average context effect was set to $\beta = 0.069$, matching the observed estimate from the early/rising model. Two additional covariates (time of day and 30-day precipitation, both scaled) were included with effect sizes matching the observed values ($\beta = -0.047$ and $\beta = -0.121$, respectively) and drawn from standard normal distributions. Age, alpha status, and reproductive

state were not explicitly simulated because the residual variance used in the data-generating process ($\sigma^2 = 0.186$) was estimated from the fitted model after conditioning on all covariates; the variance attributable to those predictors is therefore already incorporated into the noise structure of the simulation.

For each simulated dataset, I fit the same random-slope model used in the empirical analysis:

$$LOG.CORT \sim context + time\_sc + precip\_sc + (1 + context \mid ID)$$

using REML estimation with the bobyqa optimizer in lme4 (Bates et al., 2015).

The number of post-IGE samples per individual was varied across eight levels: 1, 2, 3, 5, 7, 10, 15, and 20. At each level, 500 datasets were simulated and fit. Four performance metrics were evaluated at each sample-size level:

(1) Singularity rate: the proportion of simulations in which the random slope variance was estimated at or near zero (indicating the model could not support estimation of between-individual variation in the context effect).

(2) Slope variance recovery: the median estimated random slope variance across simulations, compared to the true generating value (0.033).

(3) BLUP accuracy: the median Pearson correlation between each individual's true simulated slope and the corresponding BLUP extracted from the fitted model. This metric directly quantifies how much individual-level information the estimated component measure carries, which determines the degree of bias when BLUPs are used as predictors in downstream analyses.

(4) Power: the proportion of simulations in which the fixed effect of context (the population-average post-IGE elevation) was statistically significant at $\alpha = 0.05$.

S6.3 Results

Table S3.6. presents the simulation results across all eight sample-size levels.

Table S3.6. Simulation results: performance metrics as a function of the number of post-IGE samples per individual. Each row summarizes 500 simulated datasets. True slope variance = 0.033; true context β = 0.069; N = 70 individuals; 13 tonic samples per individual.

Post-IGE samples	% Singular	Median slope var	Median BLUP r	Power (%)
1	22.4	0.032	0.34	20.4
2	9.4	0.032	0.46	32.2
3	3.8	0.033	0.54	39.8
5	0.6	0.032	0.62	52.2
7	0.2	0.032	0.67	58.4
10	0.2	0.031	0.70	58.4
15	0.0	0.033	0.75	62.0
20	0.0	0.034	0.78	63.6

Singularity

At 1 post-IGE sample per individual (the late-window median), 22.4% of models failed to estimate a random slope. At 2 samples (the early/rising window median), singularity dropped to 9.4% but remained above the 5% threshold. Singularity fell below 5% at 3 samples per individual and was effectively eliminated (< 1%) at 5 or more samples.

Slope variance recovery.

The population-level estimate of between-individual variation was essentially unbiased at all sample sizes. The median estimated slope variance remained close to the true value (0.033) even with only 1 post-IGE sample per individual. This indicates that the mixed model can detect the presence of individual variation even when it cannot reliably attribute that variation to specific individuals.

BLUP accuracy

The correlation between true and estimated individual slopes increased monotonically with sample size but remained modest across the range examined. At 1 post-IGE sample, the median correlation was 0.34; at 2 samples, 0.46; and at 10 samples, 0.70. Even at 20 samples per individual, the correlation reached only 0.78. These values quantify the degree of measurement

error in BLUP-derived component measures: when a BLUP with an accuracy of 0.46 (the early/rising window) is used as a predictor in a downstream regression, the true coefficient is expected to be attenuated by a factor of approximately $r^2 \sim 0.21$, meaning that approximately 79% of the true association is lost to measurement error.

Power

Power to detect the population-average context effect ($\beta = 0.069$) at $\alpha = 0.05$ was 20.4% at 1 sample per individual and 32.2% at 2 samples. Power increased with additional post-IGE samples but plateaued at approximately 60–64% beyond 10 samples per individual. This plateau indicates that, for an effect of this magnitude, the number of individuals (N = 70) becomes the binding constraint on power rather than the number of post-IGE samples per individual. The observed p-value of 0.12 for the early/rising context effect is therefore consistent with a true effect of $\beta = 0.069$ sampled at a median of 2 post-IGE samples per individual, where only one in three studies would be expected to reach $p < 0.05$.

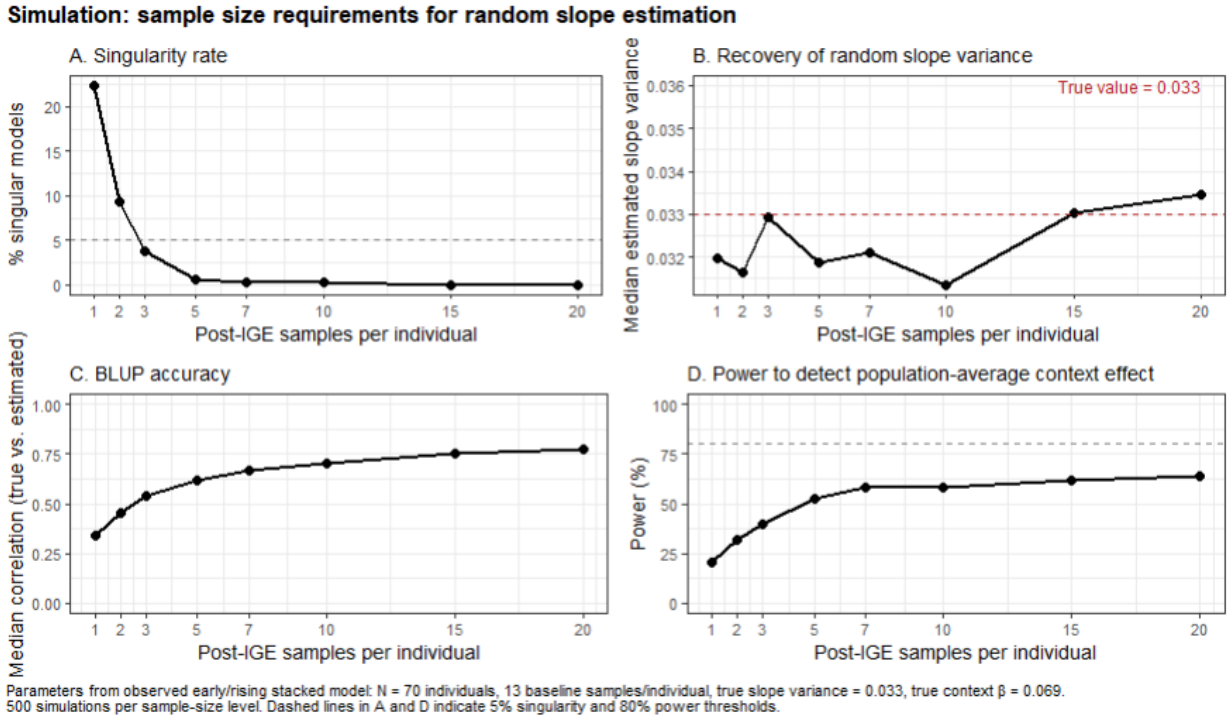


Figure S3.4. Sample size requirements for random slope estimation. Four panels show simulation-based performance metrics as a function of the number of post-IGE samples per individual, with all other parameters fixed

at observed values (N = 70 individuals, 13 tonic samples/individual, true slope variance = 0.033, true context β = 0.069). (A) Percentage of models that converged to a singular solution; dashed line indicates 5% threshold. (B) Median estimated random slope variance; dashed line indicates the true generating value. (C) Median Pearson correlation between true and estimated individual slopes (BLUP accuracy). (D) Power to detect the population-average context effect at α = 0.05; dashed line indicates 80% power.

S6.4 Implications

The observed p-value of 0.12 for the early/rising context effect should not be interpreted as evidence that intergroup encounters do not elicit a glucocorticoid response. The simulation demonstrates that with a true effect of $\beta = 0.069$ and a median of 2 post-IGE samples per individual, only 32% of studies would detect the effect at $p < 0.05$. The non-significant p-value is the expected outcome given the sampling constraints, not evidence against the hypothesis.

Further, the early/rising BLUPs extracted from the observed data carry meaningful but substantially attenuated individual-level information. At a median of 2 post-IGE samples, the estimated correlation between true and extracted slopes is approximately 0.46, indicating that the BLUPs capture less than a quarter of the true individual-level variance in reactivity ($r^2 \sim 0.21$). When these BLUPs are used as predictors in downstream analyses, this measurement error is expected to bias coefficient estimates toward zero (Houslay & Wilson, 2017). Null associations involving the early/rising responsiveness component should therefore be interpreted as reflecting limited measurement precision rather than as evidence against the hypothesized relationship.

These results provide concrete design targets for future studies. A minimum of 3 post-IGE samples per individual is needed to keep singularity below 5%, but 10 or more samples per individual are needed to achieve BLUP accuracies above 0.70. Reaching 80% power for the population-average context effect at the observed effect size would additionally require increasing the number of sampled individuals beyond 70. These targets may be achievable through study designs that prioritize serial post-stressor sampling over broader cross-sectional coverage, consistent with the recommendations in Section S3.6.

References

Altmann, J. (1974). *Observational Study of Behavior: Sampling Methods*. https://doi.org/10.1163/156853974X00534

Bailey, L. D., & van de Pol, M. (2016). Tackling extremes: challenges for ecological and evolutionary research on extreme climatic events. *Journal of Animal Ecology*, *85*(1), 85-96.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R. A., Gluckman, P., Godfrey, K., Kirkwood, T., Lahr, M. M., McNamara, J., Metcalfe, N. B., Monaghan, P., Spencer, H. G., & Sultan, S. E. (2004). Developmental plasticity and human health. *Nature*, *430*(6998), 419–421. https://doi.org/10.1038/nature02725

Baugh, A. T., Oers, K. van, Dingemanse, N. J., & Hau, M. (2014). Baseline and stress-induced glucocorticoid concentrations are not repeatable but covary within individual great tits (*Parus major*). *General and Comparative Endocrinology*, *208*, 154–163. https://doi.org/10.1016/j.ygcen.2014.08.014

Beehner, J. C., Alfaro, J., Allen, C., Benítez, M. E., Bergman, T. J., Buehler, M. S., Carrera, S. C., Chester, E. M., Deschner, T., Fuentes, A., Gault, C. M., Godoy, I., Jack, K. M., Kim, J. D., Kolinski, L., Kulick, N. K., Losch, T., Ordoñez, J. C., Perry, S. E., … Wasserman, M. D. (2022). Using an on-site laboratory for fecal steroid analysis in wild white-faced capuchins. *General and Comparative Endocrinology*, *329*, 114109. https://doi.org/10.1016/j.ygcen.2022.114109

Bell, A. M. (2005). Behavioural differences between individuals and two populations of stickleback (*Gasterosteus aculeatus*). *Journal of Evolutionary Biology*, *18*(2), 464–473. https://doi.org/10.1111/j.1420-9101.2004.00817.x

Bell, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1611), 755–761. https://doi.org/10.1098/rspb.2006.0199

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

Bergmüller, R., & Taborsky, M. (2010). Animal personality due to social niche specialisation. *Trends in Ecology & Evolution*, *25*(9), 504–511. https://doi.org/10.1016/j.tree.2010.06.012

Biro, P. A., & Stamps, J. A. (2010). Do consistent individual differences in metabolic rate promote consistent individual differences in behavior? *Trends in Ecology & Evolution*, *25*(11), 653–659. https://doi.org/10.1016/j.tree.2010.08.003

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, 9th International Symposium on Avian Endocrinology*, *163*(1), 208–213. https://doi.org/10.1016/j.ygcen.2008.12.013

Busch, D. S., & Hayward, L. S. (2009). Stress in a conservation context: A discussion of glucocorticoid actions and how levels change with conservation-relevant variables. *Biological Conservation*, *142*(12), 2844–2853. https://doi.org/10.1016/j.biocon.2009.08.013

Careau, V., Thomas, D., Humphries, M. M., & Réale, D. (2008). Energy metabolism and animal personality. *Oikos*, *117*(5), 641–653. https://doi.org/10.1111/j.0030-1299.2008.16513.x

Carrera, S. C., Godoy, I., Gault, C. M., Mensing, A., Damm, J., Perry, S. E., & Beehner, J. C. (2025). Stress responsiveness in a wild primate predicts survival across an extreme El Niño drought. *Science Advances*, *11*(4), eadq5020. https://doi.org/10.1126/sciadv.adq5020

Champagne, F. A., & Meaney, M. J. (2006). Stress During Gestation Alters Postpartum Maternal Care and the Development of the Offspring in a Rodent Model. *Biological Psychiatry*, *59*(12), 1227–1235. https://doi.org/10.1016/j.biopsych.2005.10.016

Charnov, E. L., & Berrigan, D. (1993). Why do female primates have such long lifespans and so few babies? *Or* Life in the slow lane. *Evolutionary Anthropology: Issues, News, and Reviews*, *1*(6), 191–194. https://doi.org/10.1002/evan.1360010604

Chase, I. D., Bartolomeo, C., & Dugatkin, L. A. (1994). Aggressive interactions and inter-contest interval: How long do winners keep winning? *Animal Behaviour*, 393–400.

Cockrem, J. F. (2013). Corticosterone responses and personality in birds: Individual variation and the ability to cope with environmental changes due to climate change. *General and Comparative Endocrinology, 10th International Symposium on Avian Endocrinology*, *190*, 156–163. https://doi.org/10.1016/j.ygcen.2013.02.021

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 B: Biological Sciences*, *365*(1560), 4021–4028. https://doi.org/10.1098/rstb.2010.0217

De Jong, G., & Gavrilets, S. (2000). Maintenance of genetic variation in phenotypic plasticity: The role of environmental variation. *Genetical Research*, *76*(3), 295–304. https://doi.org/10.1017/S0016672300004729

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The Adaptive Calibration Model of stress responsivity. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1562–1592. https://doi.org/10.1016/j.neubiorev.2010.11.007

Dettmer, E. L., Phillips, K. A., Rager, D. R., Bernstein, I. S., & Fragaszy, D. M. (1996). Behavioral and cortisol responses to repeated capture and venipuncture in Cebus apella. *American Journal of Primatology*, *38*(4), 357-362.

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse, N. J., Edelaar, P., & Kempenaers, B. (2010). Why is there variation in baseline glucocorticoid levels? *Trends in Ecology & Evolution*, *25*(5), 261–262. https://doi.org/10.1016/j.tree.2010.01.008

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5%253C178::AID-EVAN5%253E3.0.CO;2-8

Ellis, B. J., & Del Giudice, M. (2014). Beyond allostatic load: Rethinking the role of stress in regulating human development. *Development and Psychopathology*, *26*(1), 1–20. https://doi.org/10.1017/S0954579413000849

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*(Suppl 1).

Fedigan, L. M., Hogan, J. D., Campos, F. A., Kalbitzer, U., & Jack, K. M. (2021). Costs of male infanticide for female capuchins: When does an adaptive male reproductive strategy become costly for females and detrimental to population viability? *American Journal of Physical Anthropology*, *176*(3), 349–360. https://doi.org/10.1002/ajpa.24354

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The Complete Capuchin: The Biology of the Genus Cebus*. Cambridge University Press.

Frankenhuis, W. E., Nettle, D., & Dall, S. R. X. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 20180110. https://doi.org/10.1098/rstb.2018.0110

Gholib, G., Heistermann, M., Agil, M., Supriatna, I., Purwantara, B., Nugraha, T. P., & Engelhardt, A. (2018). Comparison of fecal preservation and extraction methods for steroid hormone metabolite analysis in wild crested macaques. *Primates*, *59*(3), 281–292. https://doi.org/10.1007/s10329-018-0653-z

Groothuis, T. G. G., & Trillmich, F. (2011). Unfolding personalities: The importance of studying ontogeny. *Developmental Psychobiology*, *53*(6), 641–655. https://doi.org/10.1002/dev.20574

Gros-Louis, J., Perry, S., & Manson, J. H. (2003). Violent coalitionary attacks and intraspecific killing in wild white-faced capuchin monkeys (Cebus capucinus). *Primates*, *44*, 341–346.

Hämäläinen, A., Heistermann, M., Fenosoa, Z. S. E., & Kraus, C. (2014). Evaluating capture stress in wild gray mouse lemurs via repeated fecal sampling: method validation and the

influence of prior experience and handling protocols on stress responses. *General and Comparative Endocrinology*, *195*, 68-79.

Harvey, P. H., & Clutton-Brock, T. H. (1985). Life History Variation in Primates. *Evolution*, *39*(3), 559–581. https://doi.org/10.1111/j.1558-5646.1985.tb00395.x

Hau, M., Casagrande, S., Ouyang, J. Q., & Baugh, A. T. (2016). Glucocorticoid-Mediated Phenotypes in Vertebrates. In *Advances in the Study of Behavior* (Vol. 48, pp. 41–115). Elsevier. https://doi.org/10.1016/bs.asb.2016.01.002

Hau, M., Deimel, C., & Moiron, M. (2022). Great tits differ in glucocorticoid plasticity in response to spring temperature. *Proceedings of the Royal Society B: Biological Sciences*, *289*.

Hau, M., & Goymann, W. (2015). Endocrine mechanisms, behavioral phenotypes and plasticity: Known relationships and open questions. *Frontiers in Zoology*, *12*(Suppl 1), S7. https://doi.org/10.1186/1742-9994-12-S1-S7

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Isler, K., & Van Schaik, C. P. (2014). How humans evolved large brains: Comparative evidence. *Evolutionary Anthropology: Issues, News, and Reviews*, *23*(2), 65–75. https://doi.org/10.1002/evan.21403

Jack, K. M., & Fedigan, L. M. (2006). Why Be Alpha Male? Dominance and Reproductive Success in Wild White-Faced Capuchins (Cebus capucinus). In A. Estrada, P. A. Garber,

M. S. M. Pavelka, & L. Luecke (Eds.), *New Perspectives in the Study of Mesoamerican Primates* (pp. 367–386). Kluwer Academic Publishers. https://doi.org/10.1007/0-387-25872-8_18

Jack, K. M., Schoof, V. A. M., Sheller, C. R., Rich, C. I., Klingelhofer, P. P., Ziegler, T. E., & Fedigan, L. (2014). Hormonal correlates of male life history stages in wild white-faced capuchin monkeys (*Cebus capucinus*). *General and Comparative Endocrinology*, *195*, 58–67. https://doi.org/10.1016/j.ygcen.2013.10.010

Kajokaite, K., Whalen, A., Panchanathan, K., & Perry, S. (2019). White-faced capuchin monkeys use both rank and relationship quality to recruit allies. *Animal Behaviour*, *154*, 161–169. https://doi.org/10.1016/j.anbehav.2019.06.008

Keller-Wood, M. E., & Dallman, M. F. (1984). Corticosteroid Inhibition of ACTH Secretion*. *Endocrine Reviews*, *5*(1), 1–24. https://doi.org/10.1210/edrv-5-1-1

Koolhaas, J. M., Korte, S. M., de Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W., & Blockhuis, H. J. (1999). Coping styles in animals: Current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, *23*, 925–935.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Lahoz, M. M., Nagle, C. A., Porta, M., Farinati, Z., & Manzur, T. D. (2007). Cortisol response and ovarian hormones in juvenile and cycling female Cebus monkeys: effect of stress and dexamethasone. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *69*(5), 551-561.

Laskowski, K. L., Chang, C.-C., Sheehy, K., & Aguiñaga, J. (2022). Consistent Individual Behavioral Variation: What Do We Know and Where Are We Going? *Annual Review of Ecology, Evolution, and Systematics*, *53*(1), 161–182. https://doi.org/10.1146/annurev-ecolsys-102220-011451

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P. M., & Meaney, M. J. (1997). Maternal Care, Hippocampal Glucocorticoid Receptors, and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science*, *277*(5332), 1659–1662. https://doi.org/10.1126/science.277.5332.1659

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Maestripieri, D., Higley, J. D., Lindell, S. G., Newman, T. K., McCormack, K. M., & Sanchez, M. M. (2006). Early maternal rejection affects the development of monoaminergic systems and adult abusive parenting in rhesus macaques (Macaca mulatta). *Behavioral Neuroscience*, *120*(5), 1017–1024. https://doi.org/10.1037/0735-7044.120.5.1017

Malkoc, K., Mentesana, L., Casagrande, S., & Hau, M. (2022). Quantifying Glucocorticoid Plasticity Using Reaction Norm Approaches: There Still is So Much to Discover! *Integrative and Comparative Biology*, *62*(1), 58–70. https://doi.org/10.1093/icb/icab196

Manson, J. H., & Perry, S. (2013). Personality structure, sex differences, and temporal change and stability in wild white-faced capuchins (Cebus capucinus). *Journal of Comparative Psychology*, *127*(3), 299–311. https://doi.org/10.1037/a0031316

McCormack, K., Newman, T. K., Higley, J. D., Maestripieri, D., & Sanchez, M. M. (2009). Serotonin transporter gene variation, infant abuse, and responsiveness to stress in rhesus

macaque mothers and infants. *Hormones and Behavior*, *55*(4), 538–547. https://doi.org/10.1016/j.yhbeh.2009.01.009

McEWEN, B. S. (1998). Stress, Adaptation, and Disease: Allostasis and Allostatic Load. *Annals of the New York Academy of Sciences*, *840*(1), 33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x

Meaney, M. J. (2001). Maternal Care, Gene Expression, and the Transmission of Individual Differences in Stress Reactivity Across Generations. *Annual Review of Neuroscience*, *24*(1), 1161–1192. https://doi.org/10.1146/annurev.neuro.24.1.1161

Mentesana, L., & Hau, M. (2022). Glucocorticoids in a warming world: Do they help birds to cope with high environmental temperatures? *Hormones and Behavior*, *142*, 105178. https://doi.org/10.1016/j.yhbeh.2022.105178

Monaghan, P. (2008). Early growth conditions, phenotypic development and environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1497), 1635–1645. https://doi.org/10.1098/rstb.2007.0011

Muniz, L., Perry, S., Manson, J. H., Gilkenson, H., Gros-Louis, J., & Vigilant, L. (2010). Male dominance and reproductive success in wild white-faced capuchins (Cebus capucinus) at Lomas Barbudal, Costa Rica. *American Journal of Primatology*, *72*(12), 1118–1130. https://doi.org/10.1002/ajp.20876

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

Nettle, D., & Bateson, M. (2015). Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151005. https://doi.org/10.1098/rspb.2015.1005

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131343. https://doi.org/10.1098/rspb.2013.1343

Palme, R. (2019). Non-invasive measurement of glucocorticoids: Advances and problems. *Physiology & Behavior*, *199*, 229–243. https://doi.org/10.1016/j.physbeh.2018.11.021

Palme, R., Rettenbacher, S., Touma, C., El-Bahr, S. M., & Möstl, E. (2005). Stress Hormones in Mammals and Birds: Comparative Aspects Regarding Metabolism, Excretion, and Noninvasive Measurement in Fecal Samples. *Annals of the New York Academy of Sciences*, *1040*(1), 162–171. https://doi.org/10.1196/annals.1327.021

Parker, K. J., & Maestripieri, D. (2011). Identifying key features of early stressful experiences that produce stress vulnerability and resilience in primates. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1466–1483. https://doi.org/10.1016/j.neubiorev.2010.09.003

Perry, S. (1996). Intergroup encounters in wild white-faced capuchins (Cebus capucinus). *International Journal of Primatology*, *17*(3).

Perry, S. (1997). Male-Female Social Relationships in Wild White-Faced Capuchins (Cebus capucinus). *Behaviour*, *134*(7/8), 477–510.

Perry, S. (1998). Male-Male Social Relationships in Wild White-Faced Capuchins, Cebus capucinus. *Behaviour*, *135*(2), 139–172.

Perry, S. (2011). Social traditions and social learning in capuchin monkeys (Cebus). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 988–996. https://doi.org/10.1098/rstb.2010.0317

Perry, S. (2012). The Behavior of Wild White-Faced Capuchins: Demography, Life History, Social Relationships, and Communication. In *Advances in the Study of Behavior* (Vol. 44, pp. 135–181). Academic Press. https://doi.org/10.1016/B978-0-12-394288-3.00004-6

Perry, S. E., Barrett, B. J., & Godoy, I. (2017). Older, sociable capuchins (*Cebus capucinus*) invent more social behaviors, but younger monkeys innovate more in other contexts. *Proceedings of the National Academy of Sciences*, *114*(30), 7806–7813. https://doi.org/10.1073/pnas.1620739114

Perry, S., Godoy, I., & Lammers, W. (2012). The Lomas Barbudal Monkey Project: Two Decades of Research on Cebus capucinus. In P. M. Kappeler & D. P. Watts (Eds.), *Long-Term Field Studies of Primates* (pp. 141–163). Springer. https://doi.org/10.1007/978-3-642-22514-7_7

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P.-O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4051–4063. https://doi.org/10.1098/rstb.2010.0208

Roff, D. (2001). *Life History Evolution* (1st ed.). Sinauer Associates.

Romero, L. M. (2004). Physiological stress in ecology: Lessons from biomedical research. *Trends in Ecology & Evolution*, *19*(5), 249–255. https://doi.org/10.1016/j.tree.2004.03.008

Romero, L. M., & Wikelski, M. (2010). Stress physiology as a predictor of survival in Galapagos marine iguanas. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1697), 3157–3162. https://doi.org/10.1098/rspb.2010.0678

Rosenberg, J., Beymer, P., Anderson, D., Van Lissa, C. j., & Schmidt, J. (2018). tidyLPA: An R Package to Easily Carry Out Latent Profile Analysis (LPA) Using Open-Source or Commercial Software. *Journal of Open Source Software*, *3*(30), 978. https://doi.org/10.21105/joss.00978

Royauté, R., & Dochtermann, N. A. (2017). When the mean no longer matters: Developmental diet affects behavioral variation but not population averages in the house cricket (*Acheta domesticus*). *Behavioral Ecology*, *28*(1), 337–345. https://doi.org/10.1093/beheco/arw164

Sanchez, M. M. (2006). The impact of early adverse care on HPA axis development: Nonhuman primate models. *Hormones and Behavior, Translational Topics in Behavioral Neuroendocrinology*, *50*(4), 623–631. https://doi.org/10.1016/j.yhbeh.2006.06.012

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions*. *Endocrine Reviews*, *21*(1), 55–89. https://doi.org/10.1210/edrv.21.1.0389

Schaebs, F. S., Perry, S. E., Cohen, D., Mundry, R., & Deschner, T. (2017). Social and demographic correlates of male androgen levels in wild white-faced capuchin monkeys (Cebus capucinus). *American Journal of Primatology*, *79*(7), e22653. https://doi.org/10.1002/ajp.22653

Schoenemann, K. L., & Bonier, F. (2018). Repeatability of glucocorticoid hormones in vertebrates: A meta-analysis. *PeerJ*, *6*, e4398. https://doi.org/10.7717/peerj.4398

Schoof, V. a. M., & Jack, K. M. (2013). The Association of Intergroup Encounters, Dominance Status, and Fecal Androgen and Glucocorticoid Profiles in Wild Male White-Faced Capuchins (Cebus capucinus). *American Journal of Primatology*, *75*(2), 107–115. https://doi.org/10.1002/ajp.22089

Sheriff, M. J., Dantzer, B., Delehanty, B., Palme, R., & Boonstra, R. (2011). Measuring stress in wildlife: Techniques for quantifying glucocorticoids. *Oecologia*, *166*(4), 869–887. https://doi.org/10.1007/s00442-011-1943-y

Sonnweber, R., Araya-Ajoy, Y. G., Behringer, V., Deschner, T., Tkaczynski, P., Fedurek, P., Preis, A., Samuni, L., Zommers, Z., Gomes, C., Zuberbühler, K., Wittig, R. M., & Crockford, C. (2018). Circadian Rhythms of Urinary Cortisol Levels Vary Between Individuals in Wild Male Chimpanzees: A Reaction Norm Approach. *Frontiers in Ecology and Evolution*, *6*. https://doi.org/10.3389/fevo.2018.00085

Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259–268. https://doi.org/10.2307/2389364

Stearns, S. C. (2000). Life history evolution: Successes, limitations, and prospects. *Naturwissenschaften*, *87*(11), 476–486. https://doi.org/10.1007/s001140050763

Stedman, J. M., Hallinger, K. K., Winkler, D. W., & Vitousek, M. N. (2017). Heritable variation in circulating glucocorticoids and endocrine flexibility in a free-living songbird. *Journal of Evolutionary Biology*, *30*(9), 1724–1735. https://doi.org/10.1111/jeb.13135

Taff, C. C., Schoenle, L. A., & Vitousek, M. N. (2018). The repeatability of glucocorticoids: A review and meta-analysis. *General and Comparative Endocrinology*, *260*, 136–145. https://doi.org/10.1016/j.ygcen.2018.01.011

Taff, C. C., & Vitousek, M. N. (2016). Endocrine Flexibility: Optimizing Phenotypes in a Dynamic World? *Trends in Ecology & Evolution*, *31*(6), 476–488. https://doi.org/10.1016/j.tree.2016.03.005

Torres-Farfan, C., Valenzuela, F. J., Ebensperger, R., Méndez, N., Campino, C., Richter, H. G., Valenzuela, G. J., & Serón-Ferré, M. (2008). Circadian cortisol secretion and circadian adrenal responses to ACTH are maintained in dexamethasone suppressed capuchin monkeys (Cebus apella). *American journal of primatology*, *70*(1), 93–100.

Touma, C., & Palme, R. (2005). Measuring Fecal Glucocorticoid Metabolites in Mammals and Birds: The Importance of Validation. *Annals of the New York Academy of Sciences*, *1046*(1), 54–74. https://doi.org/10.1196/annals.1343.006

van de Pol, M., & Cockburn, A. (2011). Identifying the critical climatic time window that affects trait expression. *The American Naturalist*, *177*(5), 698-707.

van de Pol, M., Bailey, L. D., McLean, N., Rijsdijk, L., Lawson, C. R., & Brouwer, L. (2016). Identifying the best climatic predictors in ecology and evolution. *Methods in Ecology and Evolution*, *7*(10), 1246-1257.

van Schaik, C. P., & Isler, K. (2012). Life-History Evolution in Primates. In *The Evolution of Primate Societies*. University of Chicago Press.

Vitousek, M. N., Houtz, J. L., Pipkin, M. A., Chang van Oordt, D. A., Hallinger, K. K., Uehling, J. J., Zimmer, C., & Taff, C. C. (2022). *Natural and experimental cold exposure in adulthood increase the sensitivity to future stressors in a free-living songbird*. *Functional Ecology*, 36, 2531–2543. https://doi.org/10.1111/1365-2435.14144

Weaver, I. C. g., Diorio, J., Seckl, J. R., Szyf, M., & Meaney, M. J. (2004). Early Environmental Regulation of Hippocampal Glucocorticoid Receptor Gene Expression: Characterization

of Intracellular Mediators and Potential Genomic Target Sites. *Annals of the New York Academy of Sciences*, *1024*(1), 182–212. https://doi.org/10.1196/annals.1321.099

West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford University Press.

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743. https://doi.org/10.1111/brv.12131

Wheeler, B., Tiddi, B., Kalbitzer, U., Visalberghi, E., & Heistermann, M. (2013). Methodological Considerations in the Analysis of Fecal Glucocorticoid Metabolites in Tufted Capuchins (Cebus apella). *International Journal of Primatology*, *34*, 879–898.

Wingfield, J. C. (2013). The comparative biology of environmental stress: Behavioural endocrinology and variation in ability to cope with novel, changing environments. *Animal Behaviour, Special Issue: Behavioural Plasticity and Evolution*, *85*(5), 1127–1133. https://doi.org/10.1016/j.anbehav.2013.02.018

Wingfield, J. C., Maney, D. L., Breuner, C. W., Jacobs, J. D., Lynn, S., Ramenofsky, M., & Richardson, R. D. (1998). Ecological Bases of Hormone—Behavior Interactions: The "Emergency Life History Stage." *American Zoologist*, *38*(1), 191–206. https://doi.org/10.1093/icb/38.1.191

Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). Guidelines for estimating repeatability. *Methods in Ecology and Evolution*, *3*(1), 129–137. https://doi.org/10.1111/j.2041-210X.2011.00125.x

Zimmer, C., Taff, C. C., Ardia, D. R., Ryan, T. A., Winkler, D. W., & Vitousek, M. N. (2019). On again, off again: Acute stress response and negative feedback together predict

resilience to experimental challenges. *Functional Ecology*, *33*(4), 619–628. https://doi.org/10.1111/1365-2435.13281

Chapter 4: Early-life adversity, glucocorticoid regulation, and male life-history outcomes in a wild primate population

Abstract

Early-life adversity (ELA) is hypothesized to shape adult behavior and life-history trajectories through developmental effects on physiological regulatory systems. The hypothalamic-pituitary-adrenal (HPA) axis is frequently proposed as one such mechanism, yet evidence from long-lived wild primates remains limited. Using longitudinal data from the Lomas Barbudal Monkey Project, I tested whether ELA predicts variation in adult HPA axis activity, whether HPA axis variation is associated with behavioral tendencies, and whether HPA axis and behavior explain variation in male life-history outcomes. Six forms of adversity were quantified from long-term demographic and ecological records, including rainfall deviation, maternal loss, group fission, alpha-male turnover, group size, and group mortality rate. Adult HPA axis activity was assessed from repeated fecal glucocorticoid metabolite (fGCM) measurements, distinguishing tonic activity from acute responsiveness following intergroup encounters. Behavioral analyses focused on snake-approach boldness and proactive and reactive aggression, and life-history analyses examined age at natal emigration, age at first alpha attainment, and alpha tenure duration.

Evidence for developmental programming of the HPA axis was limited. Among first-year adversity measures, only elevated mortality exposure predicted lower adult tonic glucocorticoid activity, while no adversity variable robustly predicted HPA axis reactivity after correction for multiple comparisons. In contrast, maternal loss during the first five years of life predicted

greater adult boldness and reduced behavioral flexibility, consistent with canalization of risk-taking behavior. Higher tonic glucocorticoid activity was associated with greater boldness, but neither tonic nor reactive HPA axis dimensions predicted aggression, emigration timing, alpha attainment, or alpha tenure. Male life-history outcomes were instead associated directly with developmental conditions and proactive aggression, although inference for alpha attainment was limited by the small number of attainment events. Collectively, these results provide little support for an HPA axis-mediated pathway linking early-life adversity to adult life-history variation and suggest that developmental effects in long-lived primates may often operate through behavioral and life-history channels independent of persistent glucocorticoid regulation.

Introduction

Understanding how early-life environments shape adult phenotypes remains a central problem in behavioral ecology. In long-lived, socially complex species, individuals experience substantial variation in developmental conditions, yet the pathways through which early-life adversity becomes biologically embedded and subsequently influences behavior and life-history trajectories remain incompletely understood (Monaghan, 2008; Nettle et al., 2013; Nettle & Bateson, 2015). The hypothalamic–pituitary–adrenal (HPA) axis is a plausible mechanistic link in this process (Meaney, 2001; Weaver et al., 2004; Lie et al., 1997; Champagne & Meaney, 2006). By regulating glucocorticoid production, the HPA axis mediates how organisms mobilize energy, respond to social and ecological challenges, and recover from perturbation (McEwen, 1998; Wingfield, 2013). Because glucocorticoids affect growth, immune function, metabolism, reproduction, and behavior, developmentally induced variation in HPA axis function has the potential to shape multiple components of adult phenotype (McEwen, 1998; Monaghan, 2008).

A large experimental literature demonstrates that early environments can produce lasting changes in HPA axis regulation. In rodents, variation in maternal care alters glucocorticoid receptor expression and produces stable differences in stress reactivity that persist into adulthood (Champagne & Meaney, 2006; Liu et al., 1997; Meaney, 2001; Weaver et al., 2004). Captive and semi-naturalistic primate studies likewise show that early social disruption, including maternal separation, peer rearing, and maternal abuse, can alter tonic glucocorticoid activity, acute stress responsivity, and related behavioral outcomes (Maestripieri et al., 2006; McCormack et al., 2009; Parker & Maestripieri, 2011; Sanchez, 2006). These findings establish the biological plausibility of developmental programming of the HPA axis. However, they do not resolve how naturally occurring early-life adversity affects HPA axis function in wild populations, nor how such effects should be interpreted. Altered HPA axis function could reflect developmental constraint, in which adversity imposes physiological costs that restrict later phenotype; adaptive calibration, in which early conditions provide information about future environments and development adjusts phenotype accordingly; or state-dependent adjustment, in which adversity alters an individual's somatic condition and subsequent phenotype is adjusted to that altered state without requiring early conditions to predict future environments.

This interpretive problem is especially important because the same empirical pattern can support multiple conclusions. If early adversity is associated with altered glucocorticoid regulation, increased risk-taking, accelerated reproduction, or reduced survival, one interpretation is that adversity imposes physiological costs that constrain later phenotype (McEwen, 1998; Monaghan, 2008). A second interpretation, associated with informational models of adaptive developmental plasticity, is that early-life conditions provide cues about future environments and that organisms calibrate their physiology and behavior accordingly

(Bateson et al., 2004; Del Giudice et al., 2011; Nettle & Bateson, 2015). A third possibility is that early adversity alters the individual's somatic state - its energetic reserves, physiological condition, accumulated damage, or regulatory set-points - and that later behavior and life-history decisions are made conditional on that state rather than on a forecast of the external environment (Nettle et al., 2013; Nettle & Bateson, 2015).

This distinction between informational calibration and somatic-state change is central to the present chapter. The Adaptive Calibration Model (ACM) proposes that the stress-response system is developmentally tuned to cues of environmental harshness and unpredictability, producing coordinated profiles of tonic and reactive HPA axis activity that support alternative behavioral and life-history strategies (Del Giudice et al., 2011; Ellis & Del Giudice, 2014). This framework has been influential, but its strongest assumptions are difficult to evaluate in long-lived species. For early conditions to function as predictive cues, they must provide reliable information about environments encountered much later in life. In species with extended development, long lifespans, and fluctuating social and ecological conditions, that requirement may often be weakly satisfied or absent (Fawcett & Frankenhuis, 2015; Frankenhuis et al., 2019). A somatic-state framework relaxes this assumption. From this perspective, early adversity need not forecast the future in order to shape later phenotype. Instead, the embodied consequences of adversity become part of the state on which subsequent behavior and life-history decisions depend.

HPA axis function is a useful candidate mechanism for evaluating this possibility because it can plausibly serve as both a record of developmental experience and a regulator of adult phenotype. Persistent differences in tonic glucocorticoid activity may reflect accumulated physiological wear (McEwen, 1998), altered regulatory capacity, or developmental shifts in

energetic allocation (Meaney, 2001; Weaver et al., 2004; Liu et al., 1997; Champagne & Meaney, 2006). At the same time, acute HPA axis reactivity may shape how individuals respond to immediate social and ecological challenges. In this sense, glucocorticoid regulation may provide a physiological substrate through which early-life adversity becomes embodied and later expressed in behavior and life-history trajectories (Del Giudice et al., 2011; Ellis & Del Giudice, 2014). The relevant question is therefore not simply whether adversity changes HPA axis function, but whether individual differences in HPA axis function are stable enough to be treated as features of somatic state (Nettle & Bateson, 2015), whether they are associated with behavioral phenotypes, and whether those behavioral phenotypes predict life-history outcomes.

Wild primates provide an important opportunity to evaluate these questions under conditions that complement existing human and laboratory research. In human studies, early adversity is often measured retrospectively, glucocorticoid sampling is frequently conducted in laboratory contexts, and life-history outcomes are shaped by cultural, economic, and contraceptive factors that complicate direct biological interpretation (Nettle & Bateson, 2015). Long-term primate field studies can avoid several of these limitations. Early-life conditions are observed prospectively (Weibel, 2020; Rosenbaum, 2025; Lea et al., 2015), glucocorticoids can be measured non-invasively in ecologically meaningful contexts (Carrera et al., 2025), and life-history outcomes such as dispersal, dominance attainment, reproduction, and survival can be measured directly from demographic records (Perry, 2012; Muniz et al., 2010). At the same time, wild primate studies retain the complexity that makes these questions biologically interesting: individuals develop in dynamic social groups, encounter naturally occurring stressors, and make behavioral decisions with real fitness consequences (Perry, 2012).

White-faced capuchins (*Cebus imitator*) are especially useful for this purpose. Capuchins are long-lived, socially complex, and slow-developing relative to body size, with extended juvenile periods, learned foraging skills, coalitionary social behavior, and multi-year male life-history trajectories involving natal dispersal, group entry, dominance competition, and alpha tenure (Fragaszy et al., 2004; Perry, 1998, 2011, 2012). They also provide comparative leverage because these features evolved in a platyrrhine lineage that diverged from the lineage leading to humans and apes roughly 40 million years ago. Thus, evidence that early-life adversity, HPA axis function, behavior, and life-history outcomes are linked in capuchins would strengthen the case that these processes reflect broader features of primate biology rather than patterns specific to humans or other catarrhines.

The Lomas Barbudal Monkey Project provides the longitudinal data necessary to examine these pathways in a wild population. Decades of individual-based demographic and behavioral data make it possible to characterize naturally occurring early-life adversity, adult behavior, and male life-history trajectories. Longitudinal fGCM sampling provides repeated, non-invasive measures of HPA axis activity under field conditions. In this chapter, I use these data to evaluate whether HPA axis function can be treated as an individual-level physiological phenotype, whether early-life adversity predicts adult HPA axis function, whether tonic and reactive HPA axis measures are associated with behavioral traits relevant to male reproductive competition, and whether those physiological and behavioral traits help explain variation in male life-history outcomes.

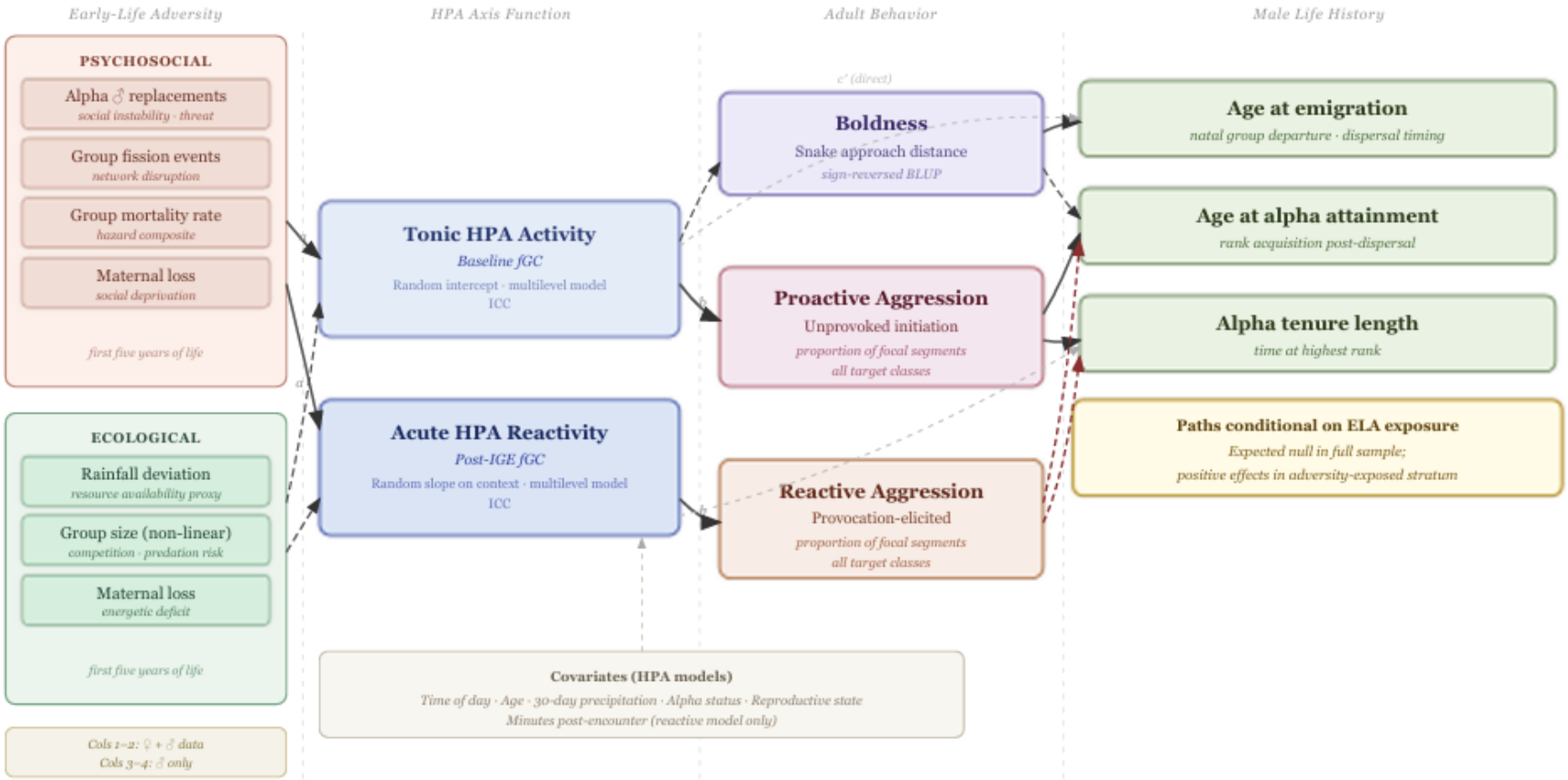
Specifically, this chapter asks three linked questions. First, does early-life adversity predict adult HPA axis function, and do different forms or developmental windows of adversity produce different physiological signatures? Second, are individual differences in HPA axis

function associated with behavioral traits such as proactive aggression, reactive aggression, and snake-response boldness? Third, do HPA axis function and behavior predict male life-history outcomes, including age at dispersal, alpha status attainment, and alpha tenure? By evaluating these links sequentially, this chapter tests whether glucocorticoid regulation plausibly functions as a somatic-state mechanism connecting early-life adversity to adult behavioral and life-history variation in a long-lived wild primate.

Methods

Figure 4.1 presents the theoretical model guiding the analyses in this chapter as a directed acyclic graph (DAG). The model proposes a four-stage causal sequence: early-life adversity (ELA) shapes individual differences in HPA axis function during development, which in turn calibrate behavioral tendencies, specifically boldness toward predators and propensity for proactive and reactive aggression, that ultimately influence the timing and success of key male life history transitions. ELA predictors are organized into two conceptually distinct clusters. Psychosocial adversity indicators (alpha male replacements, group fission events, group mortality rate, and maternal loss) index the social instability and threat exposure that the Adaptive Calibration Model (Del Giudice, Ellis, & Shirtcliff, 2011) predicts will primarily shape acute HPA axis reactivity. Ecological and energetic adversity indicators (rainfall deviation, group size, and maternal loss in its energetic dimension) index resource scarcity and physiological stress load more likely to influence chronic tonic HPA axis activity. Tonic HPA axis activity is hypothesized to predict proactive aggression, and acute HPA axis reactivity to predict reactive aggression, reflecting the neuroendocrine distinctions between instrumental and threat-elicited agonism outlined in the proactive-reactive coping style literature (Koolhaas et al., 1999). Boldness is linked primarily to emigration timing, with bolder individuals expected to tolerate the risks of natal dispersal earlier. Direct paths from HPA axis parameters to life history outcomes (shown as light dashed lines) represent residual effects not mediated through behavior.

Figure 4.1 (next page). Directed acyclic graph (DAG) depicting the theoretical model guiding analyses in this chapter. Early-life adversity (ELA) predictors are organized into psychosocial (social instability, threat) and ecological/energetic (resource availability, competition) clusters, both assessed over the first five years of life. ELA is hypothesized to shape individual differences in tonic HPA axis activity and HPA axis reactivity, which in turn calibrate adult boldness (snake approach distance BLUP) and proactive and reactive aggression. Adult behavioral tendencies are hypothesized to influence male life history outcomes. Paths from behavior to life history outcomes are expected to be conditional on ELA exposure; null associations are anticipated in the full sample. Direct paths from HPA axis parameters to life history outcomes (c') are shown as light dashed lines. Columns 1-2 include data from both sexes (ELA → HPA axis path); columns 3-4 are restricted to adult and subadult males.


Early-Life Adversity
HPA Axis Function
Adult Behavior
Male Life History
PSYCHOSOCIAL
Alpha ♂ replacements
social instability · threat
Group fission events
network disruption
Group mortality rate
hazard composite
Maternal loss
social deprivation
first five years of life
ECOLOGICAL
Rainfall deviation
resource availability proxy
Group size (non-linear)
competition · predation risk
Maternal loss
energetic deficit
first five years of life
Cols 1–2: ♀ + ♂ data
Cols 3–4: ♂ only
Tonic HPA Activity
Baseline fGC
Random intercept · multilevel model
ICC
Acute HPA Reactivity
Post-IGE fGC
Random slope on context · multilevel model
ICC
Covariates (HPA models)
Time of day · Age · 30-day precipitation · Alpha status · Reproductive state
Minutes post-encounter (reactive model only)
c′ (direct)
Boldness
Snake approach distance
sign-reversed BLUP
Proactive Aggression
Unprovoked initiation
proportion of focal segments
all target classes
Reactive Aggression
Provocation-elicited
proportion of focal segments
all target classes
Age at emigration
natal group departure · dispersal timing
Age at alpha attainment
rank acquisition post-dispersal
Alpha tenure length
time at highest rank
Paths conditional on ELA exposure
Expected null in full sample;
positive effects in adversity-exposed stratum
Primary hypothesized path
Secondary / exploratory path
Expected negative / null path
Direct effect (c′)
Covariate adjustment

Critically, the behavioral paths to life history outcomes are not expected to hold uniformly across the population. In socially complex species such as white-faced capuchins, bold and aggressive behavioral strategies are not unconditionally advantageous. Indeed, they may be costly for individuals operating in stable social environments where coalition-building and patience are more effective routes to reproductive success. Following the Adaptive Calibration Model, I predict that these behavioral associations with life history outcomes will emerge specifically among individuals who experienced significant early-life adversity, for whom a faster-paced, higher-risk behavioral strategy may represent a calibrated response to a developmental environment that signaled elevated extrinsic mortality risk and reduced time horizons. This conditionality is encoded in the DAG annotation box and motivates the exploratory adversity-exposed subsample analyses described in the statistical analysis section.

Measures

Boldness measures

Individual boldness was assessed using approach distances to dangerous snakes (boas and rattlesnakes), carried forward from the snake encounter dataset described in Chapter 2. Approach distance was log-transformed (log(distance + 0.01)) prior to analysis; on this scale, lower values indicate closer approach and thus greater boldness. Full details of the snake encounter protocol, data collection, and the repeatability estimation for snake boldness are reported in Chapter 2.

Road-crossing boldness was not included in the present analyses. Chapter 2 demonstrated that snake approach boldness and road-crossing boldness were uncorrelated at the individual level (Spearman's $\rho = -0.083$, $p = 0.565$, $N = 50$), indicating that the two assays measure distinct context-specific behavioral tendencies rather than a unified latent boldness construct. Because the snake encounter paradigm indexes risk-taking in a predator-relevant context (the ecological scenario that generates the theoretical predictions linking boldness to aggression via the spillover

hypothesis (Johnson & Sih, 2005) and the proactive coping style framework (Koolhaas et al., 1999)) it was retained as the sole boldness measure for all syndrome and flexibility analyses.

Aggression measures

Two individual-level aggression measures were derived from focal follow data collected as part of the long-term LBMP behavioral observation protocol. During focal follows, each observation segment was coded for the presence or absence of unprovoked aggressive acts (initiated by the focal individual without an immediately preceding agonistic act directed at them) and, separately, provoked aggressive acts (elicited by an agonistic act directed at the focal individual). Individual-level proactive and reactive aggression were quantified as the proportion of focal follow segments containing at least one unprovoked or provoked aggressive act, respectively, pooled across social target classes and aggregated across all available focal observations for that individual within the adult and subadult male stage. In all models predicting alpha attainment, observations made during that individual's alpha tenure were excluded. Aggression rates were logit-transformed (with a correction of 0.001 added to numerator and denominator prior to transformation) to improve distributional properties.

This operationalization maps onto the proactive–reactive coping style framework: unprovoked aggression represents calculated, instrumentally initiated conflict consistent with a proactive behavioral strategy, whereas provoked aggression represents threat-elicited escalation consistent with a reactive one. Full details of the focal follow protocol and behavioral coding procedures are reported in Chapter 2.

Male Life History Outcomes

Age at first emigration

Dispersal from the natal group is a critical life-history transition for male white-faced capuchins, who typically must emigrate before they begin reproducing (Muniz et al., 2010).

Emigration events were identified from long-term demographic records maintained by the LBMP. A male was classified as having emigrated if he disappeared from his natal group without indication of predation or illness, particularly if the disappearance was preceded by periods spent on the periphery of the group, a recent history of entering and leaving the group prior to final departure, or the simultaneous disappearance of other subadult males. For males with multiple recorded exit events from the natal group, the earliest qualifying departure date was used. Age at first emigration was computed as the years between the individuals birth date and this departure date.

For males who were recorded as having left their natal group but subsequently returned and were later recorded as immigrants into the same group, a departure was counted as a qualifying emigration event if the interval between exit and re-entry was at least seven days. This threshold reflects the distinction between exploratory forays or brief excursions during which a male tests conditions outside the natal group without committing to departure and genuine dispersal attempts, in which a male sustains independence from his natal group for a meaningful period even if he ultimately fails to establish residence elsewhere. Dispersal in male capuchins rarely proceeds directly from natal departure to successful immigration; males frequently spend extended periods moving between groups, and early failed dispersal attempts may nonetheless reflect genuine individual-level propensity to depart (Jack & Fedigan, 2006). Because the theoretical variable of interest is willingness to emigrate rather than successful resettlement, including these sustained attempts provides a more behaviorally meaningful measure of the timing of dispersal motivation and one more likely to reflect the bold, risk-tolerant phenotype that the present chapter hypothesizes to be associated with early-life adversity exposure.

Age at first alpha attainment

Alpha male status was determined from the long-term dominance records maintained by the LBMP. Male dominance rank is assigned based on the directionality of submissive behaviors, specifically "avoids" and "cowers", recorded during 10-minute focal follows and *ad libitum* observations, as these behaviors are unidirectional and yield reliable linear hierarchies (Muniz et al., 2010). For each male who attained alpha status, age at first alpha attainment was computed as the years elapsed between his birth date and the most probable start date of his earliest recorded alpha tenure, drawn from the project's alpha tenure database.

To ensure that age at first alpha attainment was known with sufficient precision, I applied a sequence of filters. First, I restricted analyses to males whose birth date was known to within +/-183 days and who survived or were observed to at least 8 years of age (the minimum age at first alpha attainment observed among males who later passed all remaining filters). Males who died or disappeared before age 8 were excluded because they are highly unlikely to have attained alpha status. For each male meeting these criteria, an end-of-follow-up date was defined as the earlier of his death date or the last date his most recent group was observed for ≥6 hours. This observation hour threshold was chosen to account for the possibility that rank dynamics are less likely to be observed on days with short observation periods. Since the planned analysis centers on rank attainment, this filter is used to minimize the probability that a change in alpha status inadvertently is overlooked by observers. Second, I excluded males for whom the combined uncertainty in age at first alpha (calculated as (age-at-first-alpha-maximum − age-at-first-alpha-minimum), which incorporates uncertainty in both the birth date and the alpha tenure start date) exceeded 183 days (approximately 6 months). Males who did not meet this criterion were not treated as alpha-attainment events but remained in the dataset as censored non-attainers at the start of the observation gap contributing uncertainty to their alpha attainment date. Third, I

identified males whose first recorded alpha tenure was preceded by an observation gap of ≥183 days occurring after age 8. Because such gaps raise the possibility that a male could have attained alpha status in an unobserved group during the interval, these males were reclassified as non-attainers and censored at the age when the gap began. One male (BD) was affected by this criterion. One additional male (CA) exceeded the combined-uncertainty threshold by a single day (184 vs. 183 days) owing to the addition of his precisely documented 4-day alpha-start window to an 180-day birth-date window. He was retained in the attainer group as a threshold case. The final analytic sample comprised 56 males, of whom 13 were classified as alpha attainers and 43 as non-attainers (including 4 known alpha males who were reclassified due to the combined-uncertainty or observation-gap criteria).

Alpha tenure length

Total alpha tenure was computed for each male who attained alpha status by summing the duration of all alpha tenures across groups. For ongoing tenures (i.e., males still holding alpha status at the end of the study period), the end date was censored at the date of the last qualifying group observation – defined as the last day on which the relevant group was observed for >6 hours. This threshold was applied because shorter observation sessions are insufficient to reliably determine alpha status. A binary censoring indicator was retained for use in survival models. Where a male held alpha status in multiple groups (e.g., following a secondary dispersal or fission event) or at multiple timepoints, tenure durations were summed across groups.

Early-life adversity

Early-life adversity measures were computed as described in Chapter 1, to which the reader is referred for full operational definitions, data sources, and descriptive statistics. Six individual-level adversity indicators were tested: rainfall harshness (deviation-weighted cumulative rainfall over the developmental window, z-scored), maternal demise (binary), group

fission (binary), alpha-male change count (z-scored), group size (z-scored), and group mortality rate (z-scored).

Two developmental windows were used. For the ELA → HPA axis analyses, the first-year window was selected a priori on the theoretical grounds that the first year of life represents a sensitive period for HPA-axis programming, during which glucocorticoid receptor expression and feedback regulation are most susceptible to environmental input in experimental studies of rodents and nonhuman primates (Meaney, 2001; Weaver et al., 2004). For all other analyses the first-five-years window was used, on the basis of Chapter 1's finding that cumulative adversity over the first five years was the most consequential developmental period for life-history outcomes, with the prenatal and first-year windows yielding few associations after correction for multiple comparisons.

Statistical Analysis

All analyses were conducted in R (version 4.3.1) using the lme4 and glmmTMB packages (Bates et al., 2015) for linear mixed-effects model fitting, with denominator degrees of freedom and p-values computed via Satterthwaite's method as implemented in lmerTest (Kuznetsova et al., 2017). Intraclass correlation coefficients (ICCs) were computed using the performance package (Lüdecke et al., 2021). All continuous predictors were scaled to a mean of zero and standard deviation of one prior to model fitting. Models that failed to converge or produced rank-deficient warnings are noted as such.

ELA → HPA axis

Effects of early-life adversity on HPA axis function were tested for the first-year developmental window, selected a priori on the theoretical grounds that the first year represents a sensitive period for HPA axis programming. Each of the six adversity indicators from Chapter 1 (rainfall deviation, maternal demise, group fission, alpha-male takeovers, group mortality rate,

and group size) was tested individually rather than as a composite. Adversity effects were estimated directly within fGCM mixed models using raw log-transformed cortisol concentrations as the outcome, rather than by correlating extracted BLUPs. This approach avoids the generated-regressor attenuation problem that affects downstream paths using BLUPs as predictors (Houslay & Wilson, 2017). As established in the rolling-window analysis (see Chapter 3), the early/rising phase of the post-encounter fGCM response is best estimated at the individual level between 0.5 and 4.5 hours post-encounter. For all reactivity analyses, post-IGE samples were therefore restricted to this window. All tonic samples were retained regardless of collection time, and post-IGE samples falling outside the 0.5–4.5-hour window were excluded.

Effects on tonic HPA axis activity were estimated from tonic samples only (i.e., samples with no confirmed intergroup encounter in the preceding 24 hours). For each adversity indicator, I fit a linear mixed model of the form:

$$\text{LOG.CORT} \sim \text{ELA} + \text{repro\_status} + \text{precip\_30d} + \text{AM.PM} + \text{age} + (1|\text{ID})$$

where ELA is the adversity indicator of interest (z-scored for continuous predictors, 0/1 for binary predictors), repro_status is a categorical covariate (male, nonpregnant female, nursing female, pregnant female), precip_30d is 30-day cumulative precipitation (z-scored), AM.PM is a binary indicator of sampling time (morning vs. afternoon/evening), age is age in years (z-scored), and (1|ID) is a random intercept for individual identity. The coefficient for ELA tests whether early-life adversity predicts adult tonic cortisol levels. Group size was tested separately as a quadratic term (gsize_c + gsize_c2) with a joint likelihood ratio test comparing models with and without both terms.

Effects on acute post-encounter responsiveness were estimated using all tonic samples and all post-IGE samples falling within the 0.5–4.5-hour window. For each adversity indicator, I fit a linear mixed model of the form:

LOG.CORT ~ ELA * post + repro_status + precip_30d + AM.PM + age + (1|ID)

where post is a binary indicator (0 = tonic, 1 = post-IGE within 0.5–4.5 hours). The interaction term ELA:post tests whether early-life adversity moderates the magnitude of the post-encounter cortisol elevation. A positive interaction indicates that higher adversity is associated with a larger post-encounter rise; a negative interaction indicates a blunted response. Group size reactivity was tested using an analogous model with interaction terms post:gsize_c and post:gsize_c2, with a joint likelihood ratio test comparing models with and without both interaction terms.

For tonic analyses, all individuals with at least one tonic sample and non-missing ELA data were included. For reactivity analyses, I required that individuals have at least one post-IGE sample within the 0.5–4.5-hour window. A sensitivity analysis restricted reactivity analyses to individuals with at least two post-IGE samples within this window to assess the robustness of findings to individual-level sampling variation.

Covariates

All models included reproductive state, 30-day cumulative precipitation, time of day (AM/PM), and age as fixed effects to account for known sources of glucocorticoid variation in this population (Carrera et al., 2025). Continuous predictors were z-scored to a mean of zero and standard deviation of one prior to analysis. Binary predictors (maternal demise, group fission) entered as 0/1. Group size was mean-centered and entered with its square to allow for non-linear effects.

ELA Effects on Boldness: Mean and Variance (Location-Scale Models)

To test whether early-life adversity shapes both the tendency and the variability of boldness responses, location-scale mixed models were fitted using glmmTMB (Brooks et al., 2017). These models jointly estimate fixed effects on the conditional mean (the location submodel) and fixed effects on the log residual standard deviation (the dispersion submodel), allowing a direct test of whether ELA alters behavioral flexibility (the within-individual variability of responses across encounters) independently of any shift in the mean level of boldness.

This approach has two key advantages over BLUP-based tests of ELA–boldness associations. First, it operates on the full trial-level dataset (N = 3,735 observations from 149 individuals with ELA data), avoiding the measurement error inherent in extracting individual-level BLUPs from a trait with low repeatability (R ~ 0.018 in Chapter 2). Second, by modeling the dispersion directly, it provides a principled test of the canalization versus anti-canalization hypotheses: a positive coefficient on the dispersion submodel indicates that ELA increases residual variability (i.e. flexibility, consistent with anti-canalization and with the hypothesis that adversity produces more flexible, environmentally responsive individuals), whereas a negative coefficient indicates that ELA reduces variability (canalization, consistent with the hypothesis that adversity narrows the behavioral phenotype and produces more rigid, consistent individuals).

The location submodel controlled for age (standardized), trial order (standardized), snake type (boa vs. cascabel), observation precision category, and sex, with a random intercept for individual identity and a sex × ELA interaction. The dispersion submodel included the ELA predictor and a random intercept for individual identity:

Location: $y \sim ELA + age\_z + trial\_order\_z + snake\_type + precision + sex + sex:ELA + (1 | monkey\_id)$

Dispersion: $\log(\sigma) \sim ELA + (1 | monkey\_id)$

To assess the robustness of any significant dispersion effects, leave-one-out (LOO) sensitivity analyses were conducted by sequentially removing each exposed individual and refitting the model, verifying that the result was not driven by any single influential case (see Supplementary Materials).

As a supplementary analysis, an unconditional flexibility model was fitted without ELA predictors in the dispersion submodel but retaining the random intercept for individual identity on log(σ). This model yields individual-specific estimates of baseline behavioral flexibility (the random effect on the dispersion intercept), representing each individual's tendency to show more or less residual variability in approach behavior than expected given the fixed-effects structure. These individual flexibility scores were extracted and standardized for use in downstream analyses.

HPA axis phenotype and Behavior

HPA axis and boldness

The association between HPA axis phenotype and boldness was estimated using a bivariate (two-response) mixed model fitted in glmmTMB (Brooks et al., 2017). Rather than reducing each individual to a single summary score and correlating those scores, this approach analyzes the two raw data streams (every individual's repeated fGCM samples and that same individual's repeated snake-approach trials) within a single model. Operationally, the two datasets are stacked into one long-format response vector, and a trait indicator labels each row as either a cortisol observation or a boldness observation. The model is best understood not as treating cortisol and approach distance as a single shared outcome, but as fitting two ordinarily

separate regressions simultaneously and linking them at the individual level. The two responses are never compared on a common scale or directly to one another; instead the model asks whether the individuals who tend to run high on cortisol are the same individuals who tend to approach snakes closely.

Trait-specific fixed effects were specified so that each covariate acted only within its own response block. Cortisol covariates included encounter context (post-IGE or tonic), time since encounter (for post-IGE samples; zero otherwise), time of day (minutes past midnight, scaled), 30-day cumulative precipitation (scaled), reproductive state (categorical), alpha status (binary), and age (scaled). Boldness covariates included age (scaled), trial order (scaled), and snake type (boa, cascabel, unknown, or other). Post-IGE samples were restricted to the 0.5–4.5 hour same-day window identified in the rolling-window analysis; samples falling outside this window were excluded entirely, as they represent neither unperturbed baseline physiology nor the biologically appropriate acute response interval. A trait-specific residual variance was estimated via the dispersion formula. Both sexes were included, with individuals required to contribute at least three snake approach trials.

The 2X2 among-individual covariance matrix, specified as correlated trait-specific random intercepts (0 + trait | ID), yields the correlation between tonic HPA axis activity (the cortisol random intercept) and mean boldness (the boldness random intercept) as a single estimated parameter, with its uncertainty propagated directly into the standard error. This joint specification was used instead of a two-step approach correlating extracted BLUPs, due to the low underlying trait repeatability associated with snake-approach boldness (see Chapter 2; Houslay & Wilson, 2017). Statistical support for the correlation was assessed by likelihood-ratio

test comparing the correlated model to one in which the two random intercepts were constrained to be independent (0 + trait || ID).

An analogous bivariate model placed the cortisol random intercept, the cortisol random slope on encounter context (acute reactivity), and the boldness random intercept in a single correlated block (0 + trait + post | ID), allowing the reactivity–boldness correlation to be estimated simultaneously with the tonic–boldness and tonic–reactivity correlations. Because the uncorrelated comparison model (using ||) constrains all three correlations to zero, the likelihood-ratio test for this model evaluates the joint set of three correlations rather than any single pair. The Wald confidence interval on each individual correlation cell was therefore used to assess which specific associations were distinguishable from zero.

HPA axis and behavioral flexibility

Associations between HPA axis activity and behavioral flexibility were assessed using a two-step approach. First, a linear mixed-effects model was fit to the endocrine data with the same covariate specification as the bivariate model above, including a random intercept and random slope on encounter context for individual identity (1 + post | ID). Individual-level BLUPs for the random intercept (tonic HPA axis) and random slope (reactivity) were extracted. Second, behavioral flexibility was computed for each individual as the within-individual standard deviation of log-transformed approach distance across snake encounters (minimum three trials). Spearman rank correlations between flexibility and each HPA axis activity were then computed. For correlations involving the reactivity BLUP, analyses were restricted to individuals contributing at least two post-IGE samples within the 0.5–4.5-hour window, as individuals with a single post-IGE sample yield slope BLUPs that are heavily shrunk toward the population mean and carry minimal individual-level information.

HPA axis and aggression

Associations between HPA axis activity and aggression were assessed using two-step correlations between individual-level HPA axis activity and individual-level aggression rates. HPA axis component measures (tonic and early/rising responsiveness) were taken from the rolling-window analysis. Individual-level aggression rates were computed from the focal follow data, restricted to non-alpha observations during the subadult and adult male periods (excluding rows with uncertain alpha status), following the operationalization described in the Measures section. For each individual, proactive aggression was quantified as the proportion of focal follow segments containing at least one unprovoked aggressive act, and reactive aggression as the proportion containing at least one provoked aggressive act, pooled across age-sex classes. Spearman rank correlations and Pearson correlations on logit-transformed rates were computed for all four pairings (tonic × proactive, tonic × reactive, rise × proactive, rise × reactive). Correlations involving the rise component measure were restricted to individuals with at least two post-IGE samples in the 0.5–4.5-hour window.

A bivariate mixed model analogous to the HPA axis–boldness specification was not feasible for aggression because aggression enters the model as binomial counts (aggressive segments out of total segments), and glmmTMB does not support mixed distributional families (Gaussian and binomial) within a single stacked model. The two-step approach is therefore used with the caveat that treating extracted BLUPs as error-free predictors is expected to attenuate correlation estimates toward zero (Houslay & Wilson, 2017). Null results involving HPA axis–aggression associations should be interpreted as potentially reflecting this conservative bias in addition to any true absence of an association.

Survival analysis of male life-history outcomes

The three male life-history outcomes (age at first emigration, age at first alpha attainment, and alpha tenure length) were analyzed using time-to-event (survival) models, which accommodate right-censored observations and use age (or for tenure the cumulative time holding alpha rank across multiple tenures) as the underlying time axis. For emigration and alpha attainment, death prior to the focal event is not independent censoring but a competing risk: an individual who dies can never subsequently emigrate or attain alpha, so treating death as ordinary censoring would overstate the probability of the event. These two outcomes were therefore analyzed in a competing-risks framework distinguishing the focal life-history transition from death.

Two complementary models were fit for each competing-risks outcome. Cause-specific Cox proportional-hazards models (survival package; Therneau, 2015) were the primary models, estimating the instantaneous rate of the focal event (emigration, alpha attainment, etc.) among individuals still at risk, and address which factors accelerate or delay the transition conditional on remaining eligible for it. Cumulative incidence functions were estimated nonparametrically using the Aalen–Johansen estimator. Fine-Gray subdistribution hazard models were fit as secondary analyses for competing-risks outcomes.

Alpha tenure length was modeled with a cause-specific Cox model only. The event is the end of an alpha tenure, and ongoing tenures (males still holding alpha status at their group's last qualifying observation) are right-censored. Death is not modeled as a separate competing risk for this outcome, so no subdistribution model was fit.

The proportional-hazards assumption was assessed for each model using scaled Schoenfeld residuals (Grambsch & Therneau, 1994), with departures noted where they occur.

Hazard ratios and 95% confidence intervals are emphasized over significance thresholds. Because the emigration analysis is well powered while the attainment and tenure analyses rest on few events, the latter two are interpreted as descriptive rather than as precise estimates of effect magnitude or direction.

ELA → male life history

For each outcome, each ELA predictor (rainfall deviation, group fission, maternal demise, alpha-male takeovers, group mortality rate, and group size) was first evaluated in a separate model. For alpha attainment, a second set of models tested whether the association between proactive aggression and attainment depended on developmental conditions by adding ELA × proactive aggression interaction terms. All were computed over the first five years of life. Continuous indicators were z-scored and binary indicators entered as 0/1; group size was centered and squared to allow for non-linearity. To test the prediction that behavioral associations with life-history outcomes are conditional on early-life adversity rather than uniform across the population, adversity x behavior interactions were fit. Each interaction was evaluated by a likelihood-ratio test comparing the additive Cox model to the model including the interaction term. Given the small number of attainment events, the robustness of any supported interaction was probed with a leave-one-out sensitivity analysis, refitting the model with each individual removed in turn and recording the resulting distribution of interaction estimates and p-values.

HPA axis → male life history

Individual HPA axis parameters were entered as predictors of each male life-history outcome using the individual-level tonic and reactivity HPA axis component measures estimated in Chapter 3. Tonic HPA axis represents each individual's estimated baseline glucocorticoid

activity, whereas reactivity HPA axis represents the magnitude of the acute glucocorticoid rise following intergroup encounters. These estimates were extracted from the kernel-window analyses described above and standardized prior to analysis.

Behavior → male life history

The primary behavioral predictors were proactive and reactive aggression. For all life-history models, aggression rates were computed exclusively from observations made during non-alpha periods as aggression expressed while holding alpha rank may be partly a consequence of rank rather than a stable antecedent of acquiring or retaining it. Restricting to non-alpha subadult- and adult-male observations yields predictors that precede, and thus are not contaminated by, the outcome. Snake-approach boldness was included as an exploratory predictor but, given its low repeatability (Chapter 2) and the small number of males with both boldness and life-history data, is not interpreted as a reliable predictor and is reported for completeness only. The conditional, adversity-moderated behavioral effects are described under the ELA → life-history analyses above.

Results

Sample Composition

The sample for Chapter 4 comprised individuals from the Lomas Barbudal population with complete data on early-life adversity measures derived from the first year and first five years of life. Sample sizes varied across analyses due to differential data availability. Sample sizes and descriptions are provided below.

Table 4.1 First-year ELA variables used in ELA→HPA axis analyses.

Variable	N	Mean	SD	Min	Max
Rainfall deviation	67	-0.202	0.402	-0.753	0.903
Group size	67	26.514	6.644	11.084	37.392
Mortality rate	67	0.039	0.062	0.000	0.263
Alpha replacements	67	0.552	1.964	0	14
Maternal loss	67	0 exposed	-	-	-

Group fission	67	2/67 exposed	-	-	-

Table 4.2 First-five-year ELA variables used in ELA→boldness analyses.

Variable	N	Mean	SD	Min	Max
Rainfall deviation	150	-0.228	0.197	-0.628	0.291
Group size	117	25.026	6.206	7.658	36.626
Mortality rate	132	0.051	0.031	0.000	0.169
Alpha replacements	150	4.307	8.355	0	31
Maternal loss	150	17/150 exposed	-	-	-
Group fission	150	76/150 exposed	-	-	-

Table 4.3 Adult behavioral and physiological phenotype

Variable	Analytic sample	N	Mean	SD	Min	Max
Snake approach distance, log-transformed	HPA-boldness overlap	94	1.986	0.173	1.500	2.429
Behavioral flexibility, SD of log-approach distance	HPA-boldness overlap	94	0.441	0.101	.168	0.894
Proactive aggression rate	HPA-aggression overlap	71	0.091	0.044	0.000	0.172
Reactive aggression rate	HPA-aggression overlap	71	0.009	0.009	0.000	0.048
Tonic HPA BLUP	HPA-boldness overlap	93	0.006	0.281	-0.775	0.543
Reactive HPA BLUP	HPA-boldness overlap	94	0.001	0.072	-0.249	0.258

Table 4.4 Male life-history outcome samples

Outcome	N	Events	Competing deaths	Censored	Mean time	SD	Range
First emigration	177	83	59	35	4.488 years	2.979	0.082-13.506
Alpha attainment	56	13	-	43	-	-	-
Alpha tenure duration	17	11 ended	-	6 ongoing	3.385 years	3.197	0.019-9.969

Early life adversity shows selective associations with adult HPA axis function.

Among the six first-year adversity indicators examined, only group mortality rate significantly predicted tonic glucocorticoid activity. Individuals exposed to higher mortality rates during their first year exhibited lower tonic cortisol concentrations in adulthood (b = -0.644, 95% CI [-1.12, -0.16], raw p = 0.009, FDR-corrected p = 0.035; Figure 4.2). No other adversity indicator-rainfall deviation, group fission, or alpha-male takeovers-showed detectable associations with tonic fGCMs (all raw p > 0.40, all FDR-corrected p = 0.679; Figure 4.2).

Maternal demise could not be evaluated because the model failed to converge due to low prevalence in the analytic sample (fewer than five exposed individuals).

Group size did not predict tonic cortisol, either linearly or quadratically (joint LRT: χ^2 = 0.171, df = 2, p = 0.918).

Glucocorticoid reactivity showed weaker and less consistent associations with early-life adversity.

Most adversity indicators were unrelated to the magnitude of post-encounter cortisol elevations. After FDR correction, no adversity indicator showed a statistically significant interaction with post-encounter status (all FDR-corrected p > 0.24). However, two marginal trends warrant transparent reporting. First, rainfall deviation showed a positive interaction with post-encounter status that approached but did not reach conventional significance (b = 0.167, 95% CI [-0.008, 0.342], raw p = 0.062, FDR-corrected p = 0.246). This effect suggests a possible trend toward larger post-encounter cortisol elevations among individuals who experienced wetter-than-average conditions during development, though the confidence interval narrowly includes zero and the effect did not survive correction for multiple comparisons. In a sensitivity analysis restricting to individuals with at least two post-encounter samples (n = 29 individuals, 96 post-encounter samples), this effect was attenuated and remained non-significant (b = 0.138, raw p = 0.153, FDR-corrected p = 0.306).

Second, alpha-male takeovers showed a marginal positive interaction in the sensitivity analysis that approached conventional significance (b = 0.096, 95% CI [-0.011, 0.203], raw p = 0.079, FDR-corrected p = 0.306), suggesting that individuals who experienced more takeovers during development may exhibit larger post-encounter cortisol elevations. However, this effect

was not present in the full sample (raw p = 0.187, FDR-corrected p = 0.373) and did not survive correction for multiple comparisons.

Group fission (full sample: b = 0.004, raw p = 0.980; sensitivity: b = 0.013, raw p = 0.934) and mortality rate (full sample: b = -0.148, raw p = 0.551; sensitivity: b = -0.134, raw p = 0.595) showed no associations with reactivity in either analysis.

Group size presented a more complex pattern. In models testing linear and quadratic effects jointly, there was no robust evidence that group size moderated post-encounter cortisol responses (joint LRT: $\chi^2 = 6.821$, df = 4, p = 0.146). A significant simple effect of the linear interaction term was observed (post:gsize_c: b = -0.039, 95% CI [-0.069, -0.009], raw p = 0.011), suggesting that individuals from larger groups may exhibit smaller post-encounter elevations. However, because the joint test-which accounts for both linear and quadratic terms simultaneously-was non-significant, this linear effect should be interpreted with caution and may reflect Type I error. The quadratic interaction term was not significant (post:gsize_c2: b = 0.002, raw p = 0.104). In the sensitivity analysis restricted to individuals with at least two post-encounter samples, the joint test remained non-significant ($\chi^2 = 7.553$, df = 4, p = 0.109).

Taken together, these analyses indicate limited programming of the HPA axis by early-life adversity in this population. The clearest and most robust signal was a long-term association between elevated first-year mortality exposure and lower tonic glucocorticoid activity-an effect that remained significant after FDR correction (FDR-corrected p = 0.035). Evidence that early social conditions shape glucocorticoid reactivity was suggestive but inconclusive: rainfall deviation and alpha-male takeovers showed marginal trends that did not survive correction for multiple comparisons, and group size showed an equivocal pattern in which a significant linear

interaction did not hold in the joint test. No other adversity indicators predicted either tonic or reactive HPA axis phenotypes.

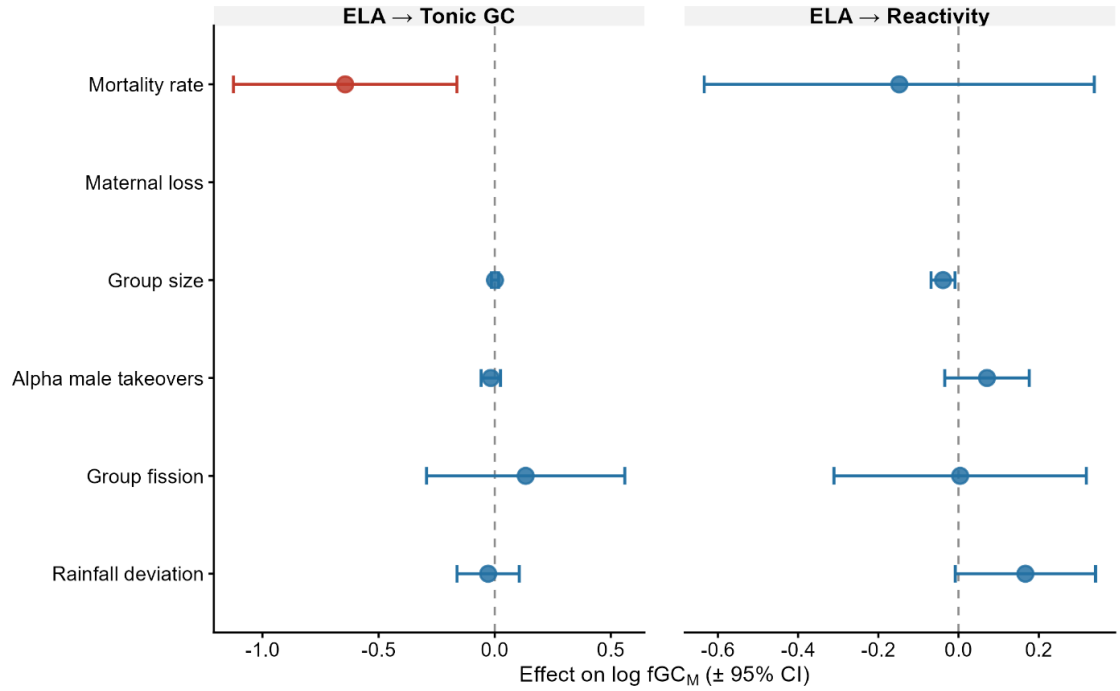


Figure 4.2. Associations between first-year early-life adversity and adult HPA axis phenotype. Points represent model coefficients (β) and horizontal bars indicate 95% confidence intervals from mixed-effects models predicting tonic glucocorticoid activity (left) and glucocorticoid reactivity following intergroup encounters (right). Tonic effects represent shifts in tonic log fGCM concentrations, whereas reactivity effects represent adversity × encounter-context interactions estimating differences in the post-encounter cortisol response. Negative coefficients indicate lower tonic activity or reduced reactivity with increasing adversity exposure. Note: Maternal loss was not estimable due to low prevalence. Red bars represent $p < 0.05$.

ELA Effects on Boldness

Mean shifts in boldness.

Location-scale models provided limited evidence that early-life adversity altered mean boldness. Maternal demise during the first five years of life was the strongest predictor of adult approach behavior. Individuals that experienced maternal loss approached snakes more closely than individuals whose mothers survived, indicating greater boldness in adulthood (b = −0.10, one-sided p = 0.032). A similar but weaker pattern was observed for mortality rate, with

individuals exposed to higher early-life mortality tending to approach more closely ($b = -0.29$, one-sided $p = 0.057$). Alpha-male turnover also showed a trend in the predicted direction ($b = -0.01$, one-sided $p = 0.087$). In contrast, rainfall, group fission, and group size showed little evidence of association with mean approach distance.

Early-life adversity and behavioral flexibility

The location-scale models revealed that different forms of adversity had distinct relationships with behavioral flexibility. The clearest effect involved maternal demise. Individuals that lost their mothers during the first five years of life exhibited significantly reduced residual variance in approach behavior ($b\sigma = -0.12$, two-tailed $p = 0.022$), indicating greater consistency across snake encounters. Thus, maternally bereaved individuals were not only bolder on average, but also less behaviorally flexible. This pattern is inconsistent with the anti-canalization hypothesis, which predicts that adversity should increase behavioral variability and produce more environmentally responsive individuals (Stamps & Groothuis, 2010). Instead, maternal loss was associated with canalization of the boldness phenotype, yielding individuals whose responses were both more risk-prone and more predictable across contexts.

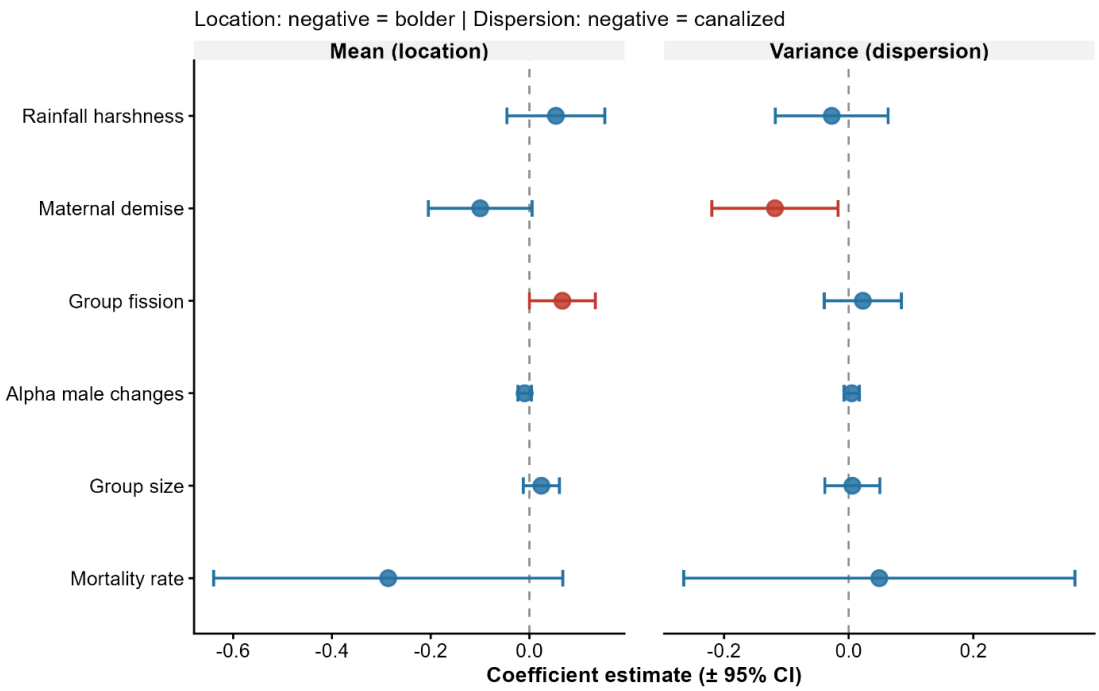


Figure 4.3. Early-life adversity predicts canalization, but not mean expression, of adult boldness. Location-scale mixed models were used to test whether six forms of early-life adversity influenced either the average level of boldness (left) or the within-individual variability of boldness responses across repeated snake encounters (right). Negative coefficients in the location model indicate greater boldness, whereas negative coefficients in the dispersion model indicate reduced behavioral variability (greater canalization). Maternal demise during the first five years of life was associated with significantly lower residual variance in boldness, indicating more consistent risk-taking behavior across contexts. No adversity measure significantly predicted mean boldness. Points show coefficient estimates ±95% confidence intervals; red bars denote $p < 0.05$.

No other adversity variable showed strong evidence of altered behavioral flexibility. Group fission, alpha-male turnover, group size, mortality rate, and rainfall harshness all produced dispersion estimates close to zero and were associated with wide confidence intervals. Although mortality rate and alpha-male turnover showed positive dispersion coefficients, neither approached conventional significance. Likewise, ecological instability measures showed little evidence of either canalization or anti-canalization.

Overall, the results indicate that early-life adversity did not produce a generalized increase in behavioral flexibility. Instead, the principal developmental signal was associated with maternal loss, which predicted a behavioral phenotype characterized by greater boldness coupled with reduced within-individual variability.

Table 4.5. Location-scale model results: ELA effects on the mean and residual variance (log σ) of snake approach distance. Positive b_sigma indicates anti-canalization (increased flexibility); negative indicates canalization (reduced flexibility). One-sided p-values test directional predictions: P(b_mean < 0) for boldness increases, P(b_sigma > 0) for anti-canalization.

ELA variable	b_mean	p (bolder)	b_sigma	p (anti-canal.)	p (two-tailed)
Rainfall harshness	+0.054	.856	-0.027	0.722	0.556
Maternal demise	-0.100	.032*	-0.118	0.989	0.022*
Group fission	+0.067	.975	+0.023	0.236	0.473
Alpha male takeovers	-0.0100	.087	+0.005	0.221	0.442
Group size	+0.024	.903	+0.006	0.398	0.796
Mortality rate	-0.286	.057	+0.049	0.379	0.759

HPA axis phenotype and Behavior

Tonic HPA axis and boldness

Higher tonic HPA axis activity was associated with greater boldness. In the bivariate model (N = 94 individuals contributing both cortisol and snake-approach data; 1,817 tonic and 159 post-IGE cortisol observations; 3,545 snake approach trials), the among-individual correlation between the cortisol random intercept and mean log approach distance was negative ($r = -0.36$, 95% CI [−0.57, −0.08]; LRT $\chi^2(1) = 6.66$, $p = 0.010$). Because lower approach distance denotes closer approach, this negative correlation indicates that individuals with higher tonic glucocorticoid concentrations approached snakes more closely-that is, behaved more boldly.

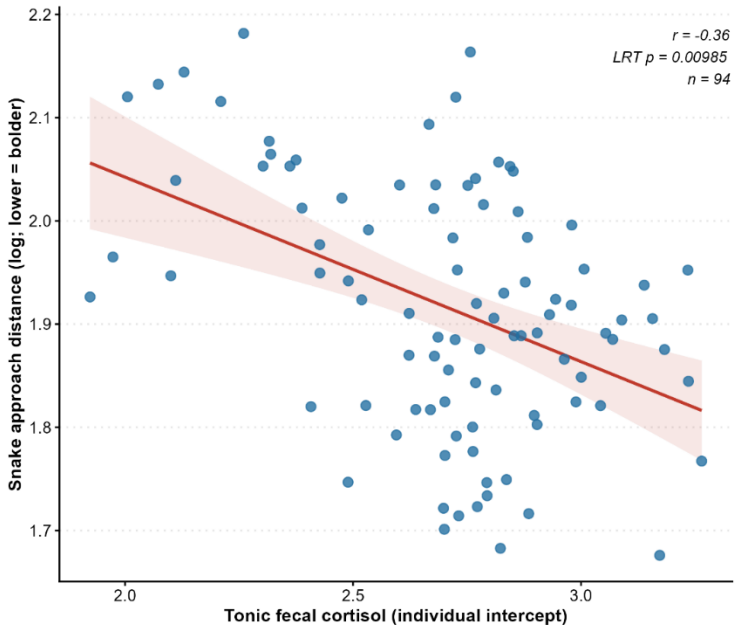


Figure 4.4. Individuals with higher tonic glucocorticoid activity are bolder. Among-individual correlation between tonic HPA axis phenotype and boldness estimated from bivariate mixed models. Boldness was quantified as the average log-transformed approach distance during repeated snake encounters, with lower values indicating closer approaches and greater risk-taking. Individuals with higher tonic glucocorticoid activity exhibited shorter approach distances and were therefore bolder ($r = -0.36$, likelihood-ratio test $p = 0.0099$, $N = 94$). Shaded region indicates the 95% confidence interval around the fitted regression line.

This result warrants a brief reconciliation with the low repeatability of snake boldness reported in Chapter 2 ($R \sim 0.02–0.09$). Low repeatability means that stable between-individual variation is small relative to total variation, not that it is absent: the bivariate model estimated a non-trivial boldness random-intercept standard deviation (0.14 on the log scale), confirming an estimable, if modest, between-individual signal. The problem low repeatability creates is specific to two-step approaches that extract a shrunken point BLUP per individual and enter it as an error-free predictor; there, low repeatability translates into noisy regressors, attenuation, and anticonservative inference. The bivariate model avoids this entirely by estimating the correlation as a covariance parameter integrated over the latent between-individual distribution rather than conditioning on point estimates. The resulting uncertainty is reflected in the confidence interval

rather than suppressed. The detection of an HPA axis–boldness correlation is therefore consistent with, not contradicted by, low boldness repeatability: a joint model can recover a true among-individual covariance from a low-repeatability trait given sufficient individuals and repeated trials, precisely because it accounts for measurement error rather than ignoring it. The point estimate should nonetheless be read together with its confidence interval, and the analysis pools both sexes.

Reactivity and boldness.

The three-component bivariate model estimated the among-individual correlation between acute HPA axis reactivity (the cortisol random slope on encounter context) and mean boldness (the boldness random intercept) at $r = -0.53$, but the 95% CI spanned zero broadly (-0.80 to $+0.35$). The joint likelihood-ratio test for all three correlations in the block was significant ($\chi^2(3) = 9.07$, $p = 0.028$), but this tests the full set of correlations simultaneously rather than isolating the reactivity–boldness cell. The Wald confidence interval for the reactivity–boldness correlation, which provides the appropriate single-parameter inference, included zero. Within this model, the tonic–reactivity correlation was near zero ($r = 0.04$, 95% CI [0.13, 0.35]), consistent with the independence of tonic and reactive HPA axis dimensions established in the rolling-window analysis. The reactivity–boldness association is therefore treated as uninterpretable on current data.

HPA axis and behavioral flexibility.

Neither HPA axis component was associated with behavioral flexibility (the within-individual standard deviation of log approach distance across snake encounters). The Spearman correlation between tonic HPA axis and flexibility showed a marginal trend that did not reach conventional significance ($\rho = -0.18$, $p = 0.080$, $N = 94$). The negative direction – higher tonic

cortisol associated with less variable approach behavior-would be consistent with the canalization pattern observed for maternal loss in the ELA–boldness analyses, but the evidence is insufficient to support that interpretation. The correlation between reactivity and flexibility, restricted to individuals with at least two post-IGE samples in the rise window, was also null (ρ = 0.21, p = 0.17, N = 43).

HPA axis and aggression.

Neither tonic HPA axis activity nor early/rising HPA axis responsiveness was associated with proactive or reactive aggression. Among the 71 males with both HPA axis component estimates and non-alpha subadult or adult aggression data, tonic HPA axis showed negligible correlations with both proactive aggression (Spearman $\rho = -0.06$, p = 0.64) and reactive aggression ($\rho = -0.10$, p = 0.43). Among the 29 males with at least two post-IGE samples in the rise window, early/rising HPA axis responsiveness was likewise unrelated to proactive aggression (ρ = 0.14, p = 0.48) and reactive aggression (ρ = 0.04, p = 0.83). Pearson correlations on logit-transformed aggression rates produced consistent nulls across all four pairings (all $|r| <$ 0.15, all $p > .33$). The DAG prediction that tonic HPA axis activity would be associated with proactive aggression, and that acute reactivity would be associated with reactive aggression, was therefore not supported.

These results should be interpreted in light of two methodological constraints. First, the HPA axis–aggression correlations use extracted BLUPs as point estimates and are therefore subject to generated-regressor attenuation, meaning null results represent conservative lower bounds rather than positive evidence against an association. Second, the rising window HPA axis analyses are restricted to a small subsample (N = 29) with limited power to detect moderate correlations. Nevertheless, the tonic–aggression correlations, which are based on a larger and

more precisely estimated sample, show effect sizes close to zero, suggesting that any true association between tonic glucocorticoid activity and aggression in this population is likely small.

HPA axis → Male Life History

Individual differences in HPA axis component did not predict any male life-history outcome. For age at first alpha attainment, neither tonic HPA axis activity nor acute HPA axis reactivity was associated with attainment timing (tonic: HR = 0.90, 95% CI [0.45, 1.80], p = 0.759; reactivity: HR = 0.91, 95% CI [0.53, 1.56], p = 0.725; N = 28, 12 attainment events). The confidence intervals for both predictors were broad and encompassed both moderate positive and negative effects, indicating little evidence that variation in glucocorticoid excretion patterns substantially influences the timing of alpha attainment. Similarly, neither HPA axis component predicted age at first emigration from the natal group. Tonic HPA axis activity was unrelated to emigration hazard (HR = 1.12, 95% CI [0.70, 1.79], p = 0.628), and acute HPA axis reactivity was likewise null (HR = 1.11, 95% CI [0.74, 1.66], p = 0.617; N = 36, 29 emigration events). Thus, males characterized by higher tonic glucocorticoid activity or stronger physiological responses to intergroup encounters did not disperse earlier or later than their peers.

The same pattern was observed for alpha tenure duration. Among males that attained alpha status, neither tonic HPA axis activity (HR = 0.92, 95% CI [0.53, 1.60], p = 0.761) nor acute HPA axis reactivity (HR = 1.36, 95% CI [0.72, 2.57], p = 0.349) predicted the hazard of tenure termination (N = 13, 10 tenure-ending events). Although the point estimate for reactivity suggested a tendency toward shorter tenures among more reactive individuals, the confidence interval was wide and included no effect. Taken together, these analyses provide little evidence that individual differences in either tonic glucocorticoid activity or physiological responsiveness

to intergroup encounters are associated with variation in male dispersal timing, alpha-rank acquisition, or alpha-tenure duration. As with all analyses using estimated individual-level HPA axis component measures as predictors, these results should be interpreted cautiously and are best viewed as evidence that any effects, if present, are likely modest relative to the uncertainty in the current dataset.

Early-Life Adversity and Male Life History Outcomes

I examined whether early-life adversity predicted three male life history outcomes: age at natal group emigration, age at first alpha attainment, and alpha tenure duration. For emigration and alpha attainment, I also tested whether adult behavioral phenotypes (boldness, proactive aggression) predicted these outcomes.

Age at Emigration

I used Cox proportional hazards models to examine the effect of early-life adversity and behavioral phenotype on age at first emigration from the natal group (N = 177, 83 emigration events; 59 deaths treated as a competing risk, 35 right-censored). Among ELA predictors, mean absolute rainfall deviation in the first five years of life was a significant predictor of emigration timing. Males who experienced greater rainfall deviation during early life had a substantially lower hazard of emigrating at any given age (HR = 0.58, 95% CI [0.40, 0.83], p = 0.003), indicating that ecologically harsher early environments were associated with delayed dispersal. Maternal loss in the first five years was also a significant predictor, with males who lost their mother emigrating earlier than those who did not (HR = 2.15, 95% CI [1.34, 3.46], p = 0.002). The remaining ELA predictors (fission exposure, alpha male instability, group size, and group mortality rate) were null (all HRs within the range 1.13–1.48, all CIs crossing 1.0).

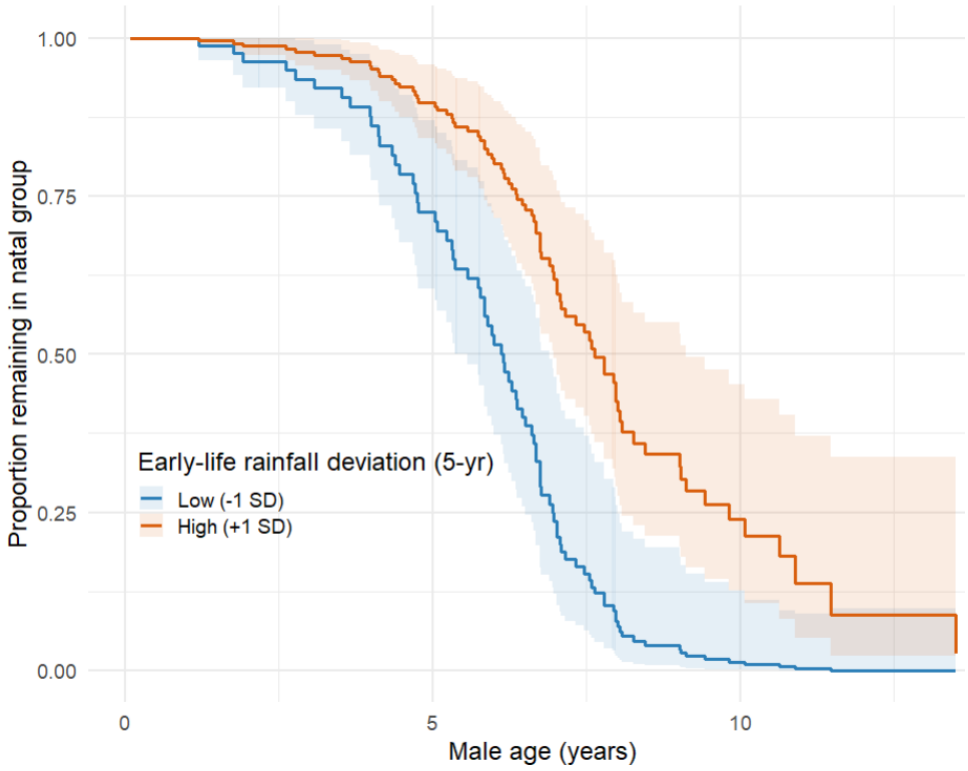


Figure 4.5. Early-life climatic perturbations is associated with delayed natal emigration. Model-predicted probability of remaining in the natal group as a function of age for males exposed to low (−1 SD; blue) and high (+1 SD; orange) levels of rainfall deviation during the first five years of life. Shaded bands indicate 95% confidence intervals. Males that experienced greater climatic instability were predicted to emigrate later than males from more stable environments, reflected in a lower hazard of emigration with increasing rainfall deviation (HR = 0.72, 95% CI: 0.56–0.91, p = 0.007).

Boldness was associated with emigration timing in the full behavioral subsample (N = 44, 27 emigration events): males with higher snake-approach boldness scores had a lower hazard of emigrating (HR = 0.63, 95% CI [0.42, 0.92]), indicating that bolder males tended to remain in their natal groups longer. While counter-intuitive from a simple boldness-as-risk-propensity perspective, this pattern may reflect the role of social context in dispersal decisions - bolder males may be better positioned to compete within their natal group, reducing the incentive to disperse. Proactive and reactive aggression rates were not associated with emigration timing (proactive: HR = 0.85, 95% CI [0.61, 1.17]; reactive: HR = 0.92, 95% CI [0.68, 1.25]), and tonic HPA axis activity showed no detectable effect in the GC subsample (HR = 1.06, 95% CI [0.68,

1.66]). Males with GC data were not systematically different in emigration age from those without (Wilcoxon p = 0.112), supporting the representativeness of that subsample. Males with aggression data, however, tended to emigrate later than those without (Wilcoxon p = 0.003), consistent with observation-derived aggression rates being conditioned on natal tenure length; the null aggression result applies specifically to males who remained long enough to become focal animals.

Age at Alpha Attainment

Across all males, the probability of alpha attainment increased steadily with age, with most attainment events occurring in early to mid-adulthood. Few males attained alpha status before approximately 9 years of age, after which the cumulative probability rose gradually.

The alpha attainment analyses were conducted on a sample of N = 56 males with birth date uncertainty ≤ 183 days, of whom 13 attained alpha status during the observation period.

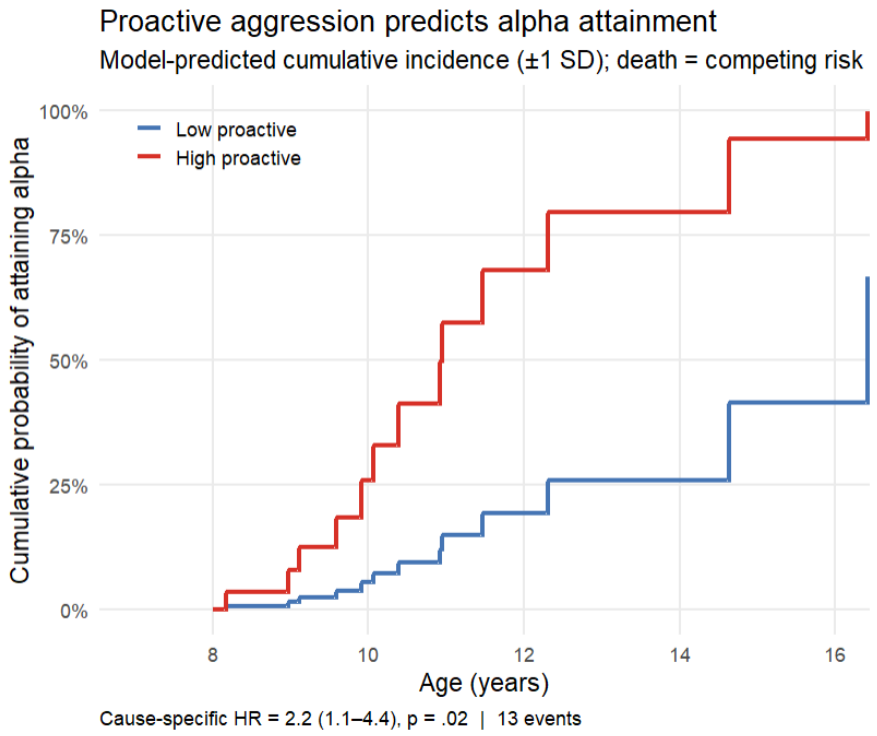


Figure 4.6. Proactive aggression is associated with increased competitive success in male capuchins. Predicted cumulative incidence of first alpha attainment from a competing-risks survival model. Red and blue curves represent males one standard deviation above and below the mean level of proactive aggression, respectively. Across the observed age range, highly proactive males were consistently more likely to achieve alpha status and reached alpha rank earlier than less proactive males. This effect was statistically significant in the cause-specific hazard model (HR = 2.2, 95% CI: 1.1–4.4, p = 0.02). Death was modeled as a competing risk. Sample sizes were 56 males and 13 alpha-attainment events.

Proactive aggression was the primary behavioral predictor of alpha attainment. Males higher in unprovoked aggression during non-alpha periods had more than twice the hazard of attaining alpha status at any given age (HR = 2.22, 95% CI [1.14, 4.35], p = 0.020; N = 48). This effect was estimated exclusively from non-alpha behavioral observations, removing the possibility that the association reflects reverse causation from rank held rather than competitive behavior prior to attainment. Reactive aggression, boldness, and tonic HPA axis activity were all null predictors of attainment timing (reactive aggression: HR = 1.20, 95% CI [0.69, 2.07]; boldness: HR = 1.73, 95% CI [0.65, 4.58]; tonic HPA axis: HR = 1.00, 95% CI [0.51, 2.00]).

Among ELA predictors, alpha male instability during the first five years of life showed a positive association with attainment hazard (HR = 1.55, 95% CI [1.08, 2.24]), suggesting that males who experienced more frequent alpha male replacements early in life may have attained rank somewhat earlier. The remaining ELA predictors were null (rainfall harshness: HR = 1.13; group size: HR = 1.19; mortality rate: HR = 1.38; fission: HR = 2.19). I additionally tested whether rainfall harshness moderated the association between proactive aggression and alpha attainment, as hypothesized under adaptive developmental plasticity frameworks. The interaction term was not significant in the current model (HR = 1.20, 95% CI [0.59, 2.43], p = 0.619; LRT p = 0.613. The ELA × proactive aggression interaction on alpha attainment is therefore null in the current dataset.

Alpha Tenure Duration

Alpha tenure analyses were conducted on N = 17 males with birth date uncertainty ≤ 183 days who attained alpha status, of whom 11 had ended tenures (event) and 6 remained alpha at last observation (right-censored). Tenure lengths ranged from 0.02 to 9.97 years. Given 11 events, all results are treated as exploratory only.

No ELA predictor was associated with tenure length (all HRs in the range 0.57–2.31, all CIs crossing 1). Among behavioral predictors, proactive aggression showed a non-significant trend toward longer tenure (HR = 0.63, 95% CI [0.37, 1.08]), consistent with a modest advantage of proactive behavioral strategies in maintaining rank once attained, though the wide confidence interval and low event count preclude firm inference. Reactive aggression was also null (HR = 1.44, 95% CI [0.75, 2.75]). Boldness showed a non-significant trend toward shorter tenure (HR = 1.86, 95% CI [0.80, 4.30]), though again this is underpowered. Tonic HPA axis activity was not associated with tenure (HR = 0.91, 95% CI [0.52, 1.58]).

A note on the interpretation of behavioral effects on alpha tenure: the aggression predictor is derived exclusively from non-alpha periods, meaning it captures pre-attainment competitive behavior rather than behavior during the tenure itself. To the extent that tenure maintenance involves different behavioral demands than rank acquisition, including coalition maintenance, affiliative behavior, and tolerance of challenges, the non-alpha aggression rate may be a poor proxy for the behavioral strategies actually deployed during the tenure. This attenuation of the proactive aggression effect from attainment (HR = 2.22) to tenure (HR = 0.63) is consistent with this framing - the competitive advantage of proactive behavior appears specific to the acquisition phase rather than persisting into tenure maintenance.

Table 4.6. Pathway summary table

Link	Result
ELA → tonic HPA axis activity	Mortality only: higher mortality → lower tonic HPA activity
ELA → reactive HPA axis activity	Weak; no robust directional effect
ELA → boldness	Maternal loss → greater boldness
ELA → flexibility	Maternal loss → reduced flexibility / greater behavioral consistency
HPA axis activity → boldness	Supported: higher tonic HPA activity → greater boldness
HPA axis activity → aggression	Unsupported
HPA axis activity → life history	Unsupported
Aggression → alpha attainment	Supported: greater proactive aggression → faster alpha attainment (note: descriptive due to low number of attainment events)

Discussion

This chapter asked whether glucocorticoid regulation functions as a somatic-state mechanism connecting early-life adversity to adult behavior and male life-history variation in a wild, long-lived primate. The answer that emerges across the four linked analyses is nuanced, modular, and much more qualified than the ACM would predict. HPA axis function does behave like a stable individual phenotype, but only in its tonic dimension; early-life adversity leaves only a faint and selective mark on the axis; adversity shapes behavior in domain-specific rather than syndrome-wide ways; and the clearest life-history consequences of adversity are direct effects on dispersal timing that point toward constraint and condition-dependence rather than accelerated life history. Taken together, the results favor developmental constraint and behavioral modularity over coordinated adaptive calibration. Only one association, between proactive aggression and rank acquisition, is consistent with, but does not confirm, an adaptive-

developmental interpretation and this finding is limited by the low number of attainment events in this dataset.

Tonic HPA axis activity is traitlike; acute reactivity is not resolvable here.

Approximately 30% of the residual variance in fGCM concentration was attributable to consistent between-individual differences (ICC = 0.295), placing tonic HPA axis activity well within the range of repeatable physiological traits documented in other wild vertebrates. Acute reactivity was a different matter. The rolling-window analysis identified a biologically plausible post-encounter signal in the 0.5–4.5-hour collection window, corresponding to a 3–4-hour gut transit time, and the early/rising stacked model estimated non-trivial random-slope variance (0.033). However, the population-average context effect did not reach conventional significance ($p = 0.12$), per-individual sampling was sparse (median = 2 post-IGE samples in the rise window), and simulation analyses indicated that the correlation between true and estimated individual slopes was only 0.46 at this sampling depth. The reactivity component is therefore retained as a secondary measure, and null results involving it should be read as inconclusive rather than as evidence that individuals do not differ in acute responsiveness.

The Adaptive Calibration Model is fundamentally a claim about coordinated profiles of tonic and reactive activity tuned to environmental cues. Because reactivity could not be estimated with confidence in this dataset, the chapter can speak to only half of that prediction. The near-zero correlation between tonic activity and early/rising responsiveness ($r = -0.07$) is at least consistent with the model's claim that the two parameters are separable dimensions, but the data cannot evaluate whether they form the integrated patterns the model envisions. Improving this would require post-encounter sampling deliberately concentrated in the early post-stressor window.

Naturally occurring adversity leaves only a faint mark on the HPA-axis.

Of the six first-year adversity indicators examined, only group mortality rate predicted tonic activity, and it did so in the direction of lower tonic glucocorticoids among individuals exposed to higher early mortality ($b = -0.644$, FDR-corrected $p = 0.035$). No other indicator showed a detectable association with tonic cortisol, and no reactivity association survived correction for multiple comparisons. This is a strikingly limited footprint relative to experimental and captive-primate studies, in which early developmental conditions have produced persistent changes in multiple components of HPA axis function. For example, variation in maternal care in rodents alters hippocampal glucocorticoid-receptor expression and subsequent HPA stress responses (Liu et al., 1997; Weaver et al., 2004), while experimental or naturally occurring disruptions of early maternal care in rhesus macaques have been associated with altered basal activity and /or cortisol and ACTH responses to stress (Sanchez, 2006; Sanchez et al., 2010; Parker & Maestripieri, 2011). However, where the literature repeatedly finds a strong effect on glucocorticoid regulation following early maternal manipulation, it should be noted here that the effect of maternal demise could not be estimated in this analysis owing to the fact that no individuals who lost their mother before age 1 made it into the analysis sample.

Two readings are compatible with the pattern. The first is the ELA → HPA axis analysis entered adversity predictors directly into the fGCM model rather than correlating them with extracted BLUPs, avoiding generated-regressor attenuation on the a-path, but the reactivity window mismatch and sparse per-individual post-IGE sampling still erode power for the reactive dimension. The second is that naturally occurring adversity in a wild population may simply not engage the developmental machinery the way severe, controlled deprivation does.

The one directional result is itself interesting. A lower tonic set-point following elevated early mortality is consistent with a blunted, hypo-cortisol profile rather than the chronically elevated baseline often invoked for chronic psychosocial stress, and it echoes the finding from this population that a lower-baseline, more responsive axis was associated with survival across an El Niño catastrophe (Carrera et al., 2025). It should be carried forward cautiously, but it is the clearest candidate for developmental shaping of the axis in these data.

Adversity shapes behavior in domain-specific ways.

The clearest behavioral signal was that maternal loss in the first five years produced individuals who were both bolder on average ($b = -0.100$, $p = 0.032$) and less variable in their responses to snakes ($b_\sigma = -0.118$, $p = 0.022$), a canalization of the boldness phenotype. No other ELA predictor produced a significant effect on either the mean or variance of boldness. Mortality rate showed a marginal trend toward greater boldness/shorter approach distance to snakes ($b = -0.286$, $p = 0.057$), a finding that converges suggestively with its status as the one ELA predictor that also predicted tonic HPA axis activity. Individuals exposed to higher early mortality may develop both a lower tonic cortisol set-point and a bolder approach phenotype, though neither the mortality–boldness association nor a formal mediational test survived conventional thresholds. Alpha-male turnover showed a marginal mean effect ($p = 0.087$) but no dispersion effect. Rainfall harshness, group fission, and group size were null on both dimensions.

The hypothesis that unpredictable developmental environments should favor retained behavioral flexibility, producing more variable, environmentally responsive individuals, received no support from any ELA predictor. The only variance-level effect detected was canalization under maternal loss, a pattern more naturally read as developmental constraint than adaptive calibration. The loss of the primary attachment figure during a sensitive period plausibly narrows

the range of behavioral strategies an individual can assemble, yielding a bold, rigid phenotype that superficially resembles the proactive coping style but arrives by a different route than the environmental-forecasting logic the calibration framework requires.

These findings reinforce the central conclusion of Chapter 2 - behavior in this cognitively complex primate is modular rather than tightly integrated into a syndrome. Adversity does not assemble a unified fast-life behavioral type. Instead, it leaves a selective, domain-specific mark on predator-response boldness, and only under a specific form of adversity (maternal loss), without propagating to social aggression or producing the coordinated covariation that pace-of-life models anticipate.

HPA axis function is linked to boldness but not to aggression, and neither propagates to life history.

Individuals with higher tonic glucocorticoid concentrations approached snakes more closely (among-individual $r = -0.36$, 95% CI [−0.57, −0.08], LRT $p = 0.010$), establishing a genuine physiology–behavior link at the level of stable individual differences and demonstrating against the backdrop of low boldness repeatability that a joint model can recover a real among-individual covariance that two-step BLUP correlations would have missed. Notably, this pairs a higher tonic axis with greater risk-taking, rather than the high-reactivity-with-vigilance coupling often emphasized in calibration profiles, another respect in which the observed structure departs from the canonical model.

That link, however, was specific to boldness. Neither tonic HPA activity nor early/rising responsiveness was associated with proactive or reactive aggression (all $|\rho| < 0.14$, all $p > .43$, $N = 29–71$). The DAG prediction that tonic HPA axis activity would channel into proactive aggression, and that acute reactivity would channel into reactive aggression, was not supported. Tonic HPA axis function thus appears coupled to a single behavioral domain without extending

to the social-competitive behaviors that most directly predict male life-history outcomes in this population. This asymmetry further reinforces the modularity of the behavioral phenotype - the physiological substrate that underlies boldness is not the same substrate that underlies aggression.

Neither tonic activity nor reactivity predicted age at emigration, age at alpha attainment, or alpha tenure length (all HRs within 0.90–1.36, all CIs spanning 1.0). Because these paths use BLUP-derived predictors, they are subject to generated-regressor attenuation and are best read as conservative lower bounds rather than positive evidence against an HPA axis–life-history link. But the broader implication for the chapter's mediation logic is clear and does not depend on the attenuation argument: with ELA → HPA axis largely null and HPA axis → life-history null, glucocorticoid regulation cannot be carrying the early-life adversity signal into the life-history outcomes that adversity does affect. Whatever routes adversity to dispersal timing in these data, it is not the measured HPA axis components.

These null findings carry a direct logical implication for the somatic-state framework that motivated this chapter. The framework proposed that HPA axis function might serve as a physiological record of developmental experience -a stable, embodied feature of individual state that both reflects early adversity and subsequently shapes behavior and life-history decisions. For this to hold, two conditions must be met: ELA must predict adult HPA axis function (the a-path), and HPA axis function must predict life-history outcomes (the b-path). Neither condition is satisfied in these data. The ELA→HPA axis analysis yielded only one association (mortality rate predicting tonic activity) that survived correction, and that association did not propagate to any life-history outcome. The HPA axis→life-history paths were uniformly null across emigration, alpha attainment, and tenure duration, with hazard ratios tightly centered around 1.0 and

confidence intervals that exclude only large effects. The somatic-state hypothesis is therefore unsupported as an account of how early adversity affects male life-history variation in this population. This represents a relatively stringent test of a pathway that has more commonly been examined in components. Studies in nonhuman primates have linked early social experience to subsequent HPA function and behavior, but tests following the full sequence from naturally occurring early-life adversity through persistent adult HPA phenotype and behavior to life-history outcomes remain rare (Brown et al., 2024; Koch et al., 2014; Lu et al., 2019; Petrullo et al., 2016). Indeed, recent work in rhesus macaques suggests that behavioral and endocrine consequences of early adversity may emerge through parallel rather than serially mediated pathways (Tromp et al., 2024). The present results similarly suggest that developmental effects on physiology, behavior, and life history likely do not constitute a single integrated causal chain.

The strongest life-history signals are direct effects on dispersal.

Males who experienced greater early-life rainfall deviation emigrated later, not earlier (HR = 0.58, p = 0.003), while males who lost their mothers emigrated earlier (HR = 2.15, p = 0.002). The rainfall result runs directly against the fast-life-history prediction that ecological adversity should accelerate the dispersal transition. It is better explained as condition-dependence or constraint where males developing under ecologically harsh conditions may be in poorer somatic state, or more risk-averse, and so defer a costly and dangerous transition. The mechanism underlying the association between maternal loss and earlier emigration is less clear. Maternal absence could alter social relationships, developmental conditions, or incentives for natal dispersal in multiple ways, and the present analyses cannot distinguish among these possibilities.

The two adversities therefore act in opposite directions on the same transition, which is itself a substantive finding. "Adversity" is not monolithic, and a single fast–slow axis fails to capture how distinct early exposures map onto dispersal. The boldness association with emigration (bolder males dispersing later, HR = 0.63) is intriguing, but it rests on the mean-boldness predictor flagged as exploratory and on a subsample conditioned on natal tenure length, and should not be over-interpreted.

Rank acquisition shows a proactive-aggression signal consistent with - but not confirming - adaptive calibration.

The one thread that aligns with an adaptive-developmental reading concerns rank. Proactive (unprovoked) aggression predicted faster alpha attainment (HR = 2.22, p = 0.020), whereas reactive aggression did not. Because the aggression measure was computed exclusively from pre-alpha observations, the association cannot be an artifact of behavior expressed once rank was held; it reflects competitive behavior preceding attainment. This is the result most compatible with the hypothesis that a proactive, instrumentally aggressive strategy pays off in male competition.

Two qualifications should be noted here. First, the analysis rests on thirteen events and is interpreted as descriptive rather than as a precise estimate. Second, and more importantly, the specific calibration prediction that the payoff to proactive aggression is conditional on early-life adversity was not supported. The rainfall harshness x proactive-aggression interaction was null (HR = 1.20, 95% CI [0.59, 2.43], p = 0.619). The data thus supports a general, unconditional advantage of proactive aggression for rank acquisition, but not the stronger claim that this advantage is a calibrated response to a harsh developmental environment.

The proactive-aggression effect also did not persist into tenure, where the coefficient attenuated to a non-significant trend in the opposite direction (HR = 0.63, 95% CI [0.37, 1.08]).

That attenuation is consistent with rank acquisition and rank maintenance making different behavioral demands. Acquisition rewards competitive aggression, whereas maintenance in this species depends heavily on coalitionary support, for which unprovoked aggression is at best irrelevant and at worst counterproductive.

Synthesis

Mapping the results onto the three interpretive frameworks set out at the start of the chapter, informational adaptive calibration is the least well supported. The coordinated tonic–reactive profiles it predicts did not materialize, reactivity could not be resolved, the imprint of adversity on the axis was faint, and the one explicitly conditional ADP prediction (adversity x proactive aggression → alpha attainment) was null. Better support exists for the developmental constraint hypothesis. Maternal loss canalizes boldness and ecological harshness delays dispersal. The somatic-state framing occupies a middle position. Tonic HPA axis activity is a stable, embodied individual trait that is coupled to boldness, but its envisioned role as a state variable mediating adversity's effect on life history is unrealized in these data, because the relevant upstream and downstream links are absent.

The larger lesson is that importing the tightly integrated fast–slow syndrome wholesale into a long-lived, cognitively complex, socially strategic primate is unwarranted. In a species where behavioral flexibility and coalitionary social structure loosen the coupling among traits, adversity does not assemble a unified high-risk phenotype, though it does leave domain-specific, often constraint-like marks, and its life-history consequences are channeled through specific transitions rather than through a coordinated physiological–behavioral strategy.

Limitations

Several limitations bear directly on interpretation here. First, every path involving an HPA axis or boldness BLUP is subject to generated-regressor attenuation, and the HPA axis–aggression correlations, which use two-step BLUP-based correlations rather than the joint bivariate model available for HPA axis–boldness, are especially susceptible. The relevant nulls are therefore conservative lower bounds rather than positive evidence of no effect. Second, the alpha attainment and tenure analyses rest on thirteen and eleven events respectively and are descriptive. Third, adult tonic activity is measured cross-sectionally, so its status as a record of developmental experience as opposed to a reflection of current state cannot be established from these data alone. The reactivity sampling-window mismatch, characterized in the rolling-window analysis, is the most addressable of these and the most consequential for testing the calibration model directly.

Conclusion

This chapter set out to test whether HPA axis function links early-life adversity to adult behavior and male life-history outcomes in wild white-faced capuchins, organized around three questions. The answers are, in brief: early-life adversity predicts adult HPA axis function only faintly and selectively; tonic HPA axis activity is associated with boldness but not with aggression, indicating a domain-specific rather than syndrome-wide physiological substrate for behavior; and HPA axis function does not predict male life-history outcomes, while early-life adversity itself acts directly on dispersal timing in a manner that points to constraint rather than accelerated life history. Proactive aggression predicts faster alpha attainment, but this advantage is unconditional on developmental experience and therefore does not confirm an adaptive-calibration interpretation.

Read against the dissertation as a whole, Chapter 4 closes the empirical arc by showing that the mechanistic pathway it was designed to trace is not realized as a connected chain in this population. Instead, the chapter recovers a stable physiological phenotype coupled to one behavioral domain but not the other and unable to mediate, a canalizing effect of maternal loss on boldness that is better explained by developmental constraint than by adaptive calibration, and direct adversity–dispersal associations that resist a single fast–slow interpretation. The one adaptive-developmental thread, that of proactive aggression predicting faster rank acquisition, is suggestive but unconditional in these data, and so remains a hypothesis about calibration rather than a demonstration of it.

The contribution is therefore more cautionary than confirmatory. It demonstrates that the adaptive-calibration and pace-of-life frameworks, which predict coordinated relationships among physiology, behavior, and life-history strategy (Réale et al., 2010; Del Giudice et al., 2011; Ellis & Del Giudice, 2014), do not transfer cleanly to a long-lived primate whose behavioral flexibility and coalitionary sociality may decouple the traits those frameworks expect to covary. Future work can sharpen the test in concrete ways: joint multivariate models that estimate HPA axis–behavior covariance directly rather than through two-stage BLUP extraction, to address the generated-regressor attenuation that limits the present analyses; accumulation of additional rank-transition events, to move the attainment and tenure analyses beyond the descriptive; and extension of the tonic HPA axis component to the female life-history outcomes examined earlier in the dissertation. Until then, the most defensible reading of these data is that early-life adversity in wild capuchins leaves real but scattered marks on physiology, behavior, and dispersal but does not assemble the coordinated, forward-looking strategy that adaptive calibration would predict.

Supplementary Information

Rainfall Sensitivity Analyses

AGE AT EMIGRATION

To assess whether the association between early-life rainfall variability and emigration timing was robust to alternative metric operationalization, I refit the Cox model using the count of 30-day windows during the first five years of life in which rainfall exceeded ±1 standard deviation from the long-term seasonal norm. Unlike the primary mean absolute deviation metric (HR = 0.70, p = 0.002), the exceedance count showed no statistically significant main effect on emigration hazard (HR = 0.83, 95% CI [0.65, 1.07], p = 0.147). Boldness again showed no main effect (HR = 1.21, 95% CI [0.76, 1.95], p = 0.421), and the interaction between rainfall exceedance and boldness was small, positive in direction, and estimated with substantial uncertainty (HR = 1.10, 95% CI [0.66, 1.84], p = 0.720). The overall model was not significant (likelihood ratio test: $\chi^2 = 2.10$, p = 0.552), confirming that the absence of boldness moderation holds across alternative operationalizations of early-life ecological variability.

Although both rainfall metrics pointed in the same direction (greater early-life rainfall variability associated with delayed emigration) the effect reached statistical significance only for the mean absolute deviation metric (HR = 0.70, p = 0.002) compared to the exceedance count metric (HR = 0.83, p = 0.147). This pattern likely reflects the greater sensitivity of the mean absolute deviation to the *magnitude* of rainfall fluctuations, which may capture cumulative

physiological load better than a simple count of threshold crossings. Consistent with this interpretation, the point estimates for the exceedance count remained in the predicted direction and its confidence interval included the primary effect estimate.

Directionality of rainfall effects

The primary emigration analyses operationalize ecological adversity as the mean absolute z-score of rainfall deviation across the first five years of life (mean_abs_z), a symmetric magnitude metric that does not distinguish between drought- and deluge-driven deviation. Because drought and deluge may have different biological consequences for a tropical dry forest primate - drought potentially reducing food availability and elevating energetic demands, deluge potentially disrupting foraging or elevating disease exposure - I conducted a series of sensitivity analyses to evaluate whether the effect of rainfall variability on emigration timing is driven primarily by one direction of deviation or reflects a symmetric response to deviation magnitude irrespective of direction.

Construction of Directional Rainfall Metrics

I constructed five additional rainfall metrics from the same underlying ERA5 daily precipitation dataset and 1940–1989 monthly z-score baseline used to derive mean_abs_z (see main text Methods and ERA5 Validation Appendix). For each individual, the first five years of life were partitioned into 60 sequential 30-day windows beginning at birth. Each window was assigned a z-score equal to the mean of the constituent monthly z-scores, weighted by the number of days falling within each calendar month. I then computed the following individual-level metrics from the 60-window series:

Signed mean z-score (mean_z_signed) is the simple arithmetic mean of all 60-window z-scores. Negative values indicate net drought conditions across the developmental window; positive values indicate net wet conditions; values near zero indicate either uniformly average conditions or balanced exposure to drought and deluge windows. This metric isolates the directional component of rainfall variability.

Drought contribution is the sum of absolute z-scores from windows with negative z-scores, divided by the total number of windows. Equivalently, it is the share of mean_abs_z that comes from below-average windows. By construction, drought_contribution + deluge_contribution = mean_abs_z, so these two metrics together additively decompose the symmetric magnitude metric into its drought and deluge components.

Deluge contribution is the analogous metric for windows with positive z-scores: the sum of absolute z-scores from above-average windows divided by total windows.

Maximum absolute z-score is the largest absolute z-score observed across the 60 windows, indicating the most extreme single deviation experienced during the developmental window in either direction.

Maximum and minimum signed z-scores are the largest positive and largest (most negative) signed z-scores observed across the 60 windows, isolating the most extreme single-window deluge and drought events, respectively.

All predictors were z-scored within the analysis dataset.

Model Specification

I fit six Cox proportional hazards models predicting age at first emigration, using the same risk set as the primary emigration analysis (N = 236, 87 emigration events, 38 deaths and 111 individuals censored at end of observation, all treated as right censored). Models differed only in their parameterization of rainfall variability:

- **Model 1 (primary):** mean_abs_z alone.
- **Model 2:** mean_z_signed alone (directional only).
- **Model 3:** drought_contribution and deluge_contribution jointly (additive directional decomposition).
- **Model 4:** mean_abs_z and mean_z_signed jointly (magnitude controlling for direction).
- **Model 5:** max_abs_z alone (single most extreme window, ignoring direction).
- **Model 6:** max_z_signed and min_z_signed jointly (most extreme deluge and drought events).

Models were compared on AIC and on the substantive interpretation of their coefficients.

Results

Table S4.1. Cox proportional hazards models comparing rainfall metric specifications for emigration timing (N = 236, 87 events).

Model	Predictor	HR	95% CI	p	AIC
1. Symmetric magnitude	mean_abs_z	0.70	[0.55, 0.88]	.002	**674.8**
2. Signed direction	mean_z_signed	0.92	[0.76, 1.10]	.358	683.9
3. Directional decomposition	drought_contribution	0.46	[0.27, 0.80]	.006	676.6
	deluge_contribution	0.44	[0.26, 0.74]	.002	
4. Magnitude + direction	mean_abs_z	0.68	[0.53, 0.88]	.003	676.6
	mean_z_signed	1.04	[0.84, 1.28]	.737	
5. Maximum absolute extreme	max_abs_z	0.81	[0.66, 1.00]	.051	680.7
6. Directional extremes	max_z_signed	0.74	[0.59, 0.93]	.011	679.6
	min_z_signed	1.24	[0.98, 1.58]	.074	

The primary symmetric magnitude metric (Model 1) yielded the lowest AIC of all models tested. The directional decomposition (Model 3) showed that drought contribution and deluge contribution had similar effect estimates in the same direction (HRs of 0.46 and 0.44 respectively, $p = 0.006$ and $p = 0.002$), with greater deviation in either direction associated with later emigration. The magnitudes of these coefficients differ from the symmetric model because

each represents the effect of variation in only one half of the deviation distribution; together they sum to the symmetric metric.

The signed-direction-only model (Model 2) was null ($p = 0.36$), with the worst AIC of all models tested. When signed direction was added to the symmetric magnitude model (Model 4), it contributed no additional explanatory power (signed mean $p = 0.74$), and AIC increased by 1.9 relative to Model 1. The symmetric magnitude metric retained its full effect after adjustment for signed direction ($HR = 0.68$, $p = 0.003$).

The two extreme-event specifications produced more equivocal results. The maximum absolute z-score (Model 5) was a marginal predictor ($HR = 0.81$, $p = 0.051$) with worse fit than the cumulative metrics. The directional extremes model (Model 6) showed an asymmetric pattern: the wettest single window predicted delayed emigration ($HR = 0.74$, $p = 0.011$), while the driest single window showed a weak effect in the opposite direction ($HR = 1.24$, $p = 0.074$). This pattern is opposite in direction to the cumulative decomposition result and is confined to single-event extremes; both extreme-event models had poorer fit (higher AIC) than models based on cumulative across-window deviation.

Interpretation

The convergence of evidence across models supports a symmetric-magnitude interpretation of the rainfall effect on emigration timing. Both drought and deluge magnitudes contribute to delayed emigration with effects that are statistically indistinguishable in size, and the directional component of rainfall variability (signed mean) does not predict emigration either alone or after adjusting for magnitude. Models that isolate the cumulative magnitude of deviation across the entire developmental window fit better than models based on single most extreme events, suggesting that the effect reflects chronic exposure to deviating conditions rather than acute response to extreme events.

The asymmetric pattern observed in the directional-extremes model (Model 6) where the wettest single window delays emigration but the driest single window weakly accelerates it is interesting should be interpreted cautiously given the marginal significance of the drought-side coefficient and the worse overall model fit. The dominant interpretation supported by these analyses is that *deviation from historical rainfall norms in either direction* during the first five years of life predicts delayed natal emigration, consistent with frameworks emphasizing developmental responses to environmental instability rather than to specific harshness cues. This finding aligns more closely with predictions from unpredictability-focused extensions of the Adaptive Calibration Model and adaptive developmental plasticity frameworks (Del Giudice et al., 2011; Nettle & Bateson, 2015) than with classic harshness-driven life history models that predict directional effects of resource scarcity.

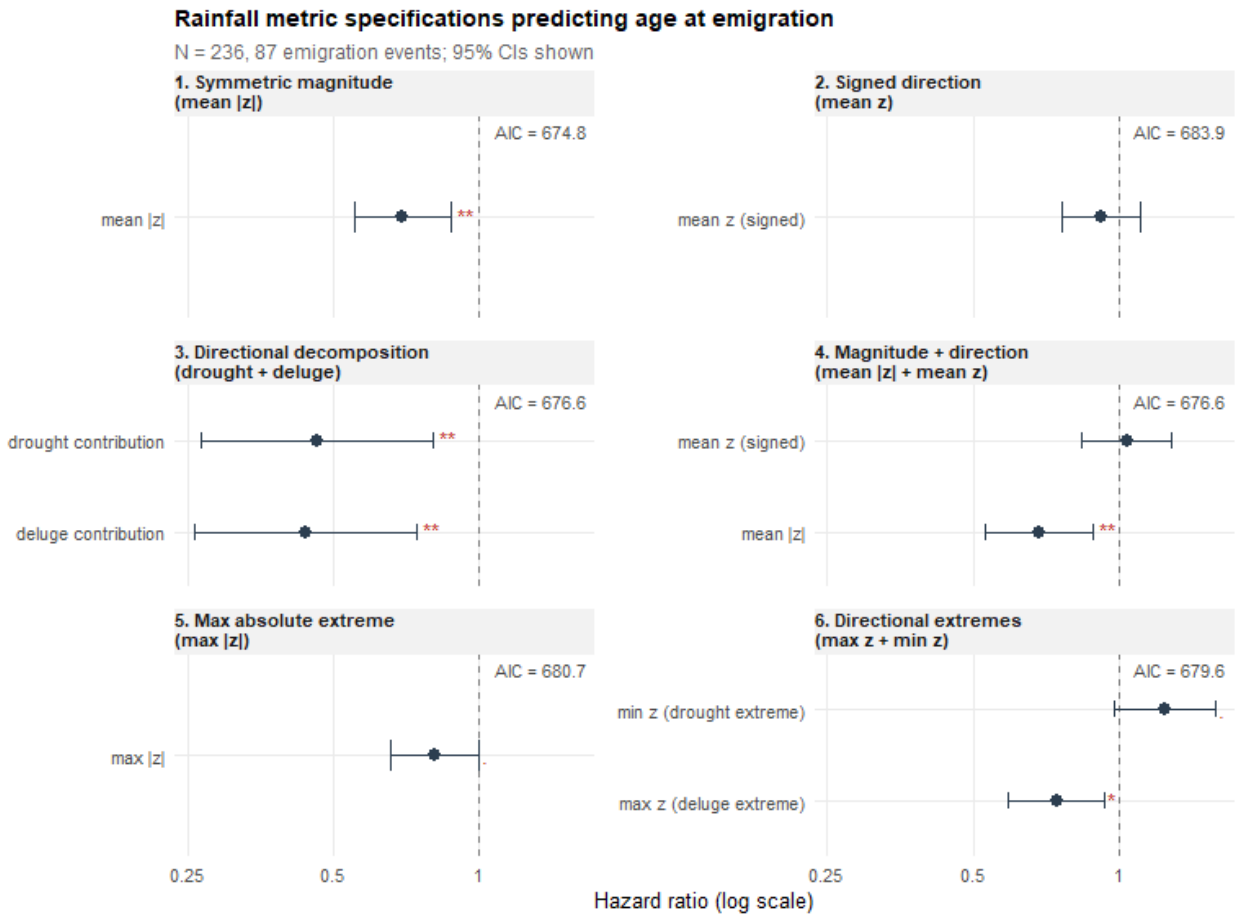


Figure S4.1. Rainfall variability and age at emigration. Cox proportional hazards models comparing alternative parameterizations of early-life rainfall variability (first five years of life). Points show hazard ratios (log scale) with 95% confidence intervals. Models isolating the cumulative magnitude of deviation from seasonal norms (mean |z|) provide the best fit (lowest AIC), while models based on directional effects (mean z) or single extreme events perform worse.

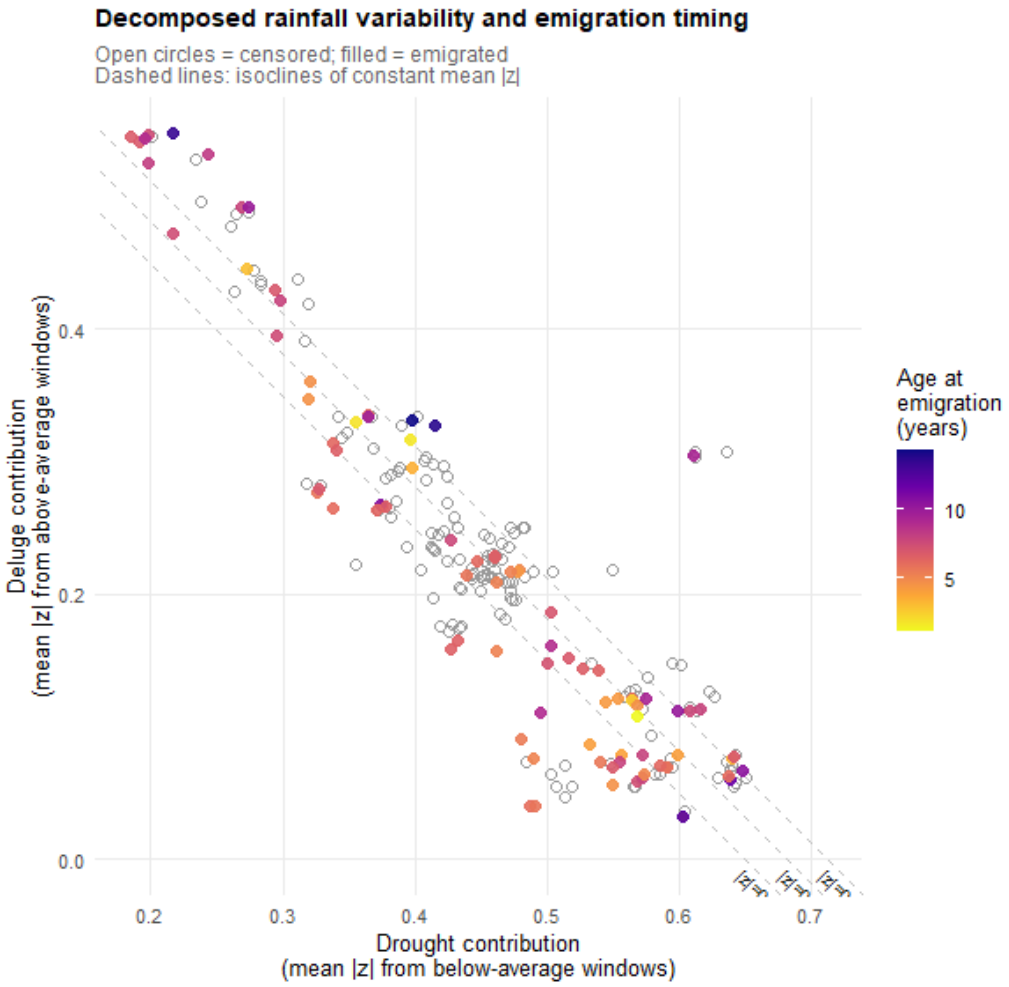


Figure S4.2. Decomposition of rainfall variability into drought (x-axis) and deluge (y-axis) contributions. Each point represents an individual; filled circles indicate emigrants and open circles indicate censored observations. Color denotes age at emigration (years). Dotted lines indicate isoclines of constant total rainfall variability (mean |z|), such that movement across lines reflects increasing total variability, while movement along a line reflects changes in the balance of drought versus deluge at constant variability. Variation in emigration timing aligns with total variability rather than directional composition, indicating that cumulative magnitude of rainfall deviation-irrespective of whether driven by drought or deluge-predicts delayed emigration.

Age at alpha attainment

Using an alternative measure of ecological variability based on the number of months during the first five years of life in which precipitation exceeded ±1 standard deviation from the long-term mean, I again found no evidence of a main effect of rainfall exceedance on alpha attainment (HR = 0.73, $p = 0.30$). Proactive aggression remained positively associated with earlier alpha attainment (HR = 3.23, $p = 0.019$). The interaction between rainfall exceedance and proactive aggression was positive in direction and of moderate magnitude (HR = 2.96), although

estimated with uncertainty ($p = 0.075$). The overall model including both predictors and their interaction was significant (likelihood ratio test: $p = 0.026$), indicating that this pattern is consistent across alternative operationalizations of early-life ecological variability.

Sensitivity Analyses Across Developmental Windows (HPA Axis)

To evaluate whether associations between early-life adversity and HPA axis phenotype were specific to the first-year developmental window, I repeated the tonic and reactivity analyses using analogous prenatal and first-five-year adversity measures. These analyses were conducted using the same modeling framework as the primary analyses and are presented as sensitivity analyses rather than independent hypothesis tests.

For tonic glucocorticoid activity, no prenatal or first-five-year adversity measure was significantly associated with baseline cortisol concentrations after correction for multiple comparisons (Table S4.2). In particular, the association between first-year mortality rate and lower tonic glucocorticoid activity observed in the primary analyses did not generalize to either late prenatal mortality exposure ($\beta = -0.023$, $p = 0.587$) or first-five-year mortality exposure ($\beta = -0.673$, $p = 0.226$). Thus, the mortality-rate effect appeared specific to the first-year developmental window.

Sensitivity analyses of glucocorticoid reactivity yielded a similarly selective pattern. No evidence was found for effects of rainfall deviation, group fission, maternal demise, mortality rate, or alpha-male takeovers in either the prenatal or first-five-year models (all FDR-adjusted $p > 0.05$). Group size was the only adversity dimension to show recurring associations with reactivity across developmental windows. In the prenatal analyses, larger early-life group size was associated with a nonsignificant tendency toward reduced post-encounter GC responses ($\beta =$

-0.021, p = 0.147), and the joint likelihood-ratio test for linear and quadratic group-size interactions approached significance ($\chi^2 = 5.39$, df = 2, p = 0.067). In the first-five-year analyses, larger group size predicted reduced GC reactivity (β = -0.057, p = 0.0018; FDR-adjusted p = 0.018), and the joint likelihood-ratio test supported inclusion of the group-size interaction terms ($\chi^2 = 6.56$, df = 2, p = 0.038).

Taken together, these sensitivity analyses suggest that the tonic glucocorticoid effect of mortality exposure is restricted to the first year of life, whereas the relationship between early-life group size and glucocorticoid reactivity may reflect broader developmental social conditions spanning multiple early-life periods. However, because these analyses were conducted as secondary sensitivity tests, they should be interpreted cautiously.

Table S4.2. HPA axis activity sensitivity analyses across developmental windows. P-values were adjusted separately within tonic and reactivity model families using the Benjamini-Hochberg false discovery rate procedure.

Window	Outcome	Predictor	β	95% CI	p	FDR-adjusted p
First 5 Years	Reactivity	Group fission	0.191	-0.224, 0.606	0.368	0.919
First 5 Years	Reactivity	Group size	-0.057	-0.093, -0.021	0.0018	0.0176
First 5 Years	Reactivity	Maternal demise	-0.068	-0.712, 0.577	0.837	0.986
First 5 Years	Reactivity	Mortality rate	0.746	-1.590, 3.080	0.531	0.986
First 5 Years	Reactivity	Rainfall deviation	-0.037	-0.432, 0.358	0.855	0.986
First 5 Years	Reactivity	Alpha takeovers	-0.003	-0.067, 0.061	0.930	0.986
First 5 Years	Tonic	Group fission	0.009	-0.172, 0.191	0.921	0.921
First 5 Years	Tonic	Group size	0.007	-0.008, 0.022	0.389	0.890
First 5 Years	Tonic	Maternal demise	-0.119	-0.369, 0.132	0.352	0.890
First 5 Years	Tonic	Mortality rate	-0.673	-1.760, 0.415	0.226	0.890
First 5 Years	Tonic	Rainfall deviation	0.065	-0.126, 0.256	0.503	0.890
First 5 Years	Tonic	Alpha takeovers	-0.013	-0.051, 0.025	0.493	0.890
Prenatal	Reactivity	Group fission	0.009	-1.030, 1.050	0.986	0.986
Prenatal	Reactivity	Group size	-0.021	-0.049, 0.007	0.147	0.489

Prenatal	Reactivity	Mortality rate	-	-	-	-
Prenatal	Reactivity	Rainfall deviation	0.061	-0.274, 0.396	0.721	0.986
Prenatal	Reactivity	Alpha takeovers	-0.173	-0.400, 0.055	0.137	0.489
Prenatal	Tonic	Group fission	-0.112	-0.731, 0.507	0.723	0.890
Prenatal	Tonic	Group size	0.002	-0.010, 0.014	0.728	0.890
Prenatal	Tonic	Mortality rate	-0.023	-0.106, 0.060	0.587	0.890
Prenatal	Tonic	Rainfall deviation	0.087	-0.050, 0.224	0.212	0.890
Prenatal	Tonic	Alpha takeovers	0.008	-0.135, 0.150	0.917	0.921

Table S4.3. Joint likelihood-ratio tests for linear and quadratic group-size effects.

Window	Outcome	χ^2	df	p
First 5 Years	Reactivity	6.56	2	0.0377
First 5 Years	Tonic	3.07	2	0.215
Prenatal	Reactivity	5.39	2	0.0674
Prenatal	Tonic	1.22	2	0.544

References

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R. A., Gluckman, P., Godfrey, K., Kirkwood, T., Lahr, M. M., McNamara, J., Metcalfe, N. B., Monaghan, P., Spencer, H. G., & Sultan, S. E. (2004). Developmental plasticity and human health. *Nature*, *430*(6998), 419–421. https://doi.org/10.1038/nature02725

Brown, E. R., Gettler, L. T., & Rosenbaum, S. (2024). Effects of social environments on male primate HPG and HPA axis developmental programming. *Developmental psychobiology*, *66*(5), e22491. https://doi.org/10.1002/dev.22491

Champagne, F. A., & Meaney, M. J. (2006). Stress During Gestation Alters Postpartum Maternal Care and the Development of the Offspring in a Rodent Model. *Biological Psychiatry*, *59*(12), 1227–1235. https://doi.org/10.1016/j.biopsych.2005.10.016

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The Adaptive Calibration Model of stress responsivity. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1562–1592. https://doi.org/10.1016/j.neubiorev.2010.11.007

Ellis, B. J., & Del Giudice, M. (2014). Beyond allostatic load: Rethinking the role of stress in regulating human development. *Development and Psychopathology*, *26*(1), 1–20. https://doi.org/10.1017/S0954579413000849

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*(Suppl 1).

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The Complete Capuchin: The Biology of the Genus Cebus*. Cambridge University Press.

Frankenhuis, W. E., Nettle, D., & Dall, S. R. X. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 20180110. https://doi.org/10.1098/rstb.2018.0110

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Jack, K. M., & Fedigan, L. M. (2006). Why Be Alpha Male? Dominance and Reproductive Success in Wild White-Faced Capuchins (Cebus capucinus). In A. Estrada, P. A. Garber, M. S. M. Pavelka, & L. Luecke (Eds.), *New Perspectives in the Study of Mesoamerican Primates* (pp. 367–386). Kluwer Academic Publishers. https://doi.org/10.1007/0-387-25872-8_18

Koch, H., McCormack, K., Sanchez, M. M., & Maestripieri, D. (2014). The development of the hypothalamic-pituitary-adrenal axis in rhesus monkeys: effects of age, sex, and early experience. *Developmental psychobiology*, *56*(1), 86–95. https://doi.org/10.1002/dev.21093

Lea, A. J., Altmann, J., Alberts, S. C., & Tung, J. (2015). Developmental Constraints in a Wild Primate. *The American Naturalist*, *185*(6), 809–821. https://doi.org/10.1086/681016

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P. M., & Meaney, M. J. (1997). Maternal Care, Hippocampal Glucocorticoid Receptors, and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science*, *277*(5332), 1659–1662. https://doi.org/10.1126/science.277.5332.1659

Lu, A., Petrullo, L., Carrera, S., Feder, J., Schneider-Crease, I., & Snyder-Mackler, N. (2019). Developmental responses to early-life adversity: Evolutionary and mechanistic perspectives. *Evolutionary anthropology*, *28*(5), 249–266. https://doi.org/10.1002/evan.21791

Maestripieri, D., Higley, J. D., Lindell, S. G., Newman, T. K., McCormack, K. M., & Sanchez, M. M. (2006). Early maternal rejection affects the development of monoaminergic systems and adult abusive parenting in rhesus macaques (Macaca mulatta). *Behavioral Neuroscience*, *120*(5), 1017–1024. https://doi.org/10.1037/0735-7044.120.5.1017

McCormack, K., Newman, T. K., Higley, J. D., Maestripieri, D., & Sanchez, M. M. (2009). Serotonin transporter gene variation, infant abuse, and responsiveness to stress in rhesus macaque mothers and infants. *Hormones and Behavior*, *55*(4), 538–547. https://doi.org/10.1016/j.yhbeh.2009.01.009

McEWEN, B. S. (1998). Stress, Adaptation, and Disease: Allostasis and Allostatic Load. *Annals of the New York Academy of Sciences*, *840*(1), 33–44. https://doi.org/10.1111/j.1749-6632.1998.tb09546.x

Meaney, M. J. (2001). Maternal Care, Gene Expression, and the Transmission of Individual Differences in Stress Reactivity Across Generations. *Annual Review of Neuroscience*, *24*(1), 1161–1192. https://doi.org/10.1146/annurev.neuro.24.1.1161

Monaghan, P. (2008). Early growth conditions, phenotypic development and environmental change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1497), 1635–1645. https://doi.org/10.1098/rstb.2007.0011

Muniz, L., Perry, S., Manson, J. H., Gilkenson, H., Gros-Louis, J., & Vigilant, L. (2010). Male dominance and reproductive success in wild white-faced capuchins (Cebus capucinus) at

Lomas Barbudal, Costa Rica. *American Journal of Primatology*, *72*(12), 1118–1130. https://doi.org/10.1002/ajp.20876

Nettle, D., & Bateson, M. (2015). Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151005. https://doi.org/10.1098/rspb.2015.1005

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131343. https://doi.org/10.1098/rspb.2013.1343

Parker, K. J., & Maestripieri, D. (2011). Identifying key features of early stressful experiences that produce stress vulnerability and resilience in primates. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1466–1483. https://doi.org/10.1016/j.neubiorev.2010.09.003

Perry, S. (1998). Male-Male Social Relationships in Wild White-Faced Capuchins, Cebus capucinus. *Behaviour*, *135*(2), 139–172.

Perry, S. (2011). Social traditions and social learning in capuchin monkeys (Cebus). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1567), 988–996. https://doi.org/10.1098/rstb.2010.0317

Perry, S. (2012). The Behavior of Wild White-Faced Capuchins: Demography, Life History, Social Relationships, and Communication. In *Advances in the Study of Behavior* (Vol. 44, pp. 135–181). Academic Press. https://doi.org/10.1016/B978-0-12-394288-3.00004-6

Petrullo, L. A., Mandalaywala, T. M., Parker, K. J., Maestripieri, D., & Higham, J. P. (2016). Effects of early life adversity on cortisol/salivary alpha-amylase symmetry in free-ranging juvenile rhesus macaques. *Hormones and Behavior*, *86*, 78-84.

Rosenbaum, S., Malani, A., Lea, A. J., Tung, J., Alberts, S. C., & Archie, E. A. (2025). Testing early life effects frameworks: developmental constraints and adaptive response hypotheses do not explain fertility outcomes in wild female baboons. *Proceedings of the Royal Society B: Biological Sciences*, *292*(2050), 20242485.

Sanchez, M. M. (2006). The impact of early adverse care on HPA axis development: Nonhuman primate models. *Hormones and Behavior, Translational Topics in Behavioral Neuroendocrinology*, *50*(4), 623–631. https://doi.org/10.1016/j.yhbeh.2006.06.012

Sanchez, M. M., Mccormack, K., Grand, A. P., Fulks, R., Graff, A., & Maestripieri, D. (2010). Effects of sex and early maternal abuse on adrenocorticotropin hormone and cortisol responses to the corticotropin-releasing hormone challenge during the first 3 years of life in group-living rhesus monkeys. *Development and psychopathology*, *22*(1), 45-53.

Stamps, J., & Groothuis, T. G. (2010). The development of animal personality: relevance, concepts and perspectives. *Biological Reviews*, *85*(2), 301-325.

Tromp, D. P. M., Fox, A. S., Riedel, M. K., Oler, J. A., Zhou, X., Roseboom, P. H., Alexander, A. L., & Kalin, N. H. (2024). Early life adversity in primates: Behavioral, endocrine, and neural effects. *Psychoneuroendocrinology*, *162*, 106953. https://doi.org/10.1016/j.psyneuen.2023.106953

Weaver, I. C. g., Diorio, J., Seckl, J. R., Szyf, M., & Meaney, M. J. (2004). Early Environmental Regulation of Hippocampal Glucocorticoid Receptor Gene Expression: Characterization of Intracellular Mediators and Potential Genomic Target Sites. *Annals of the New York Academy of Sciences*, *1024*(1), 182–212. https://doi.org/10.1196/annals.1321.099

Weibel, C. J., Tung, J., Alberts, S. C., & Archie, E. A. (2020). Accelerated reproduction is not an adaptive response to early-life adversity in wild baboons. *Proceedings of the National Academy of Sciences*, *117*(40), 24909–24919. https://doi.org/10.1073/pnas.2004018117

Wingfield, J. C. (2013). The comparative biology of environmental stress: Behavioural endocrinology and variation in ability to cope with novel, changing environments. Animal Behaviour, Special Issue: Behavioural Plasticity and Evolution, 85(5), 1127–1133. https://doi.org/10.1016/j.anbehav.2013.02.018

Conclusion

This dissertation asks, "does early-life adversity 'get under the skin' in wild, long-lived, socially complex primates - and if so, through what mechanisms"? The organizing hypothesis was a four-stage causal chain: naturally occurring adversity encountered during development programs adult hypothalamic–pituitary–adrenal (HPA) function, which in turn shapes behavioral phenotypes, which in turn shape fitness-relevant life history outcomes. Each link in this chain has theoretical support from work in laboratory rodents, captive primates, and humans, and the chain as a whole has become a popular framework in developmental evolutionary psychology. The capuchin case explored here is informative precisely because it provides an opportunity to evaluate all four links simultaneously, in naturalistic conditions, across individuals whose developmental histories are known prospectively. As a result, we able to ask not just whether the chain holds, but where and why it breaks.

The answer that emerges from more than three decades of individual-based data on white-faced capuchins (*Cebus imitator*) at Lomas Barbudal is instructive, but nuanced. The full hypothesized pathway from ELA → HPA axis function → behavior → life history strategies did not hold. Early adversity did not consistently program HPA axis function; HPA axis function did not reliably predict behavioral phenotypes across domains; behavioral phenotypes did not propagate uniformly to life history outcomes; and the few significant links that emerged did not form a coherent mediated pathway from developmental experience through physiology and behavior to fitness. But this pattern of results is not a null finding in the sense of a series of underpowered non-detections, although power was limited for particular variables and links in the hypothesized pathway. It is instead a specific pattern of failures concentrated at points where the ACM framework makes its most demanding assumptions, specifically (a) that early environmental

cues reliably forecast adult conditions in a long-lived species; (b) that a single physiological axis integrates developmental experience into a coordinated behavioral profile; and (c) that population-level behavioral syndromes are an appropriate target for detecting individual developmental effects. Instead, I find that (a) early adversity is lethal, and viability selection removes the most affected before adulthood; (b) psychosocial adversity may influence reproductive timing among surviving females; (c) maternal loss canalizes boldness; (d) proactive aggression predicts alpha attainment; and (e) tonic and reactive HPA axis are independent – not as a set of failed hypotheses but as a coherent set of findings about how developmental experience, physiology, and behavior each independently contribute to fitness outcomes along sex-specific pathways.

I. Summary of Empirical Findings

Chapter 1: Early-Life Adversity and Female Life History

Chapter 1 examined whether naturally occurring early-life adversity predicted female mortality risk, the timing of reproductive milestones, and reproductive output across three developmental windows: the late prenatal period, the first year of life, and the first five years of life. The analysis tested the competing predictions of predictive adaptive response (PAR) models, which expect adversity to accelerate reproductive timing, against developmental constraint models, which expect adversity to delay or suppress reproduction (Nettle & Bateson, 2015; Nettle et al., 2013).

The primary delayed-entry survival models, which conditioned on female survival to age five, yielded no statistically significant associations between early-life adversity and subsequent mortality risk in any developmental window. This null result was universal: no individual predictor reached significance across the prenatal, first-year, or first-five-year models. However, the

complementary from-birth sensitivity analyses, which removed the left-truncation requirement and followed all females from birth (n = 187, 107 deaths), revealed that early-life adversity is in fact incredibly lethal when experienced in the prenatal period or across the first-year of life. Prenatal exposure to alpha male replacement quadrupled mortality risk (HR = 4.34, p = 0.015). Maternal loss during the first year produced a comparable elevation (HR = 4.43, p = 0.002).

The contrast between the from-birth and delayed-entry results provides direct evidence for viability selection: early-life adversity kills during the juvenile period, and by conditioning on survival to age five, the delayed-entry specification removes the most adversity-affected individuals from the risk set. In other words, the most adversity-exposed females are dead before they can display any programming effect on adult survival or reproduction. This is not an absence of adversity effects but a signature of selection operating before the observational window opens.

The primary models for age at first birth (AFB), using the 4.5-year biological minimum for reproduction, were uniformly null across all three developmental windows. However, a sensitivity analysis using a six-year reproductive threshold (restricting the sample to females who had plausibly entered the period of reproductive readiness) revealed that psychosocial adversity during the first five years of life significantly predicted earlier first birth. Each additional alpha male replacement was associated with a 7.2% increase in the rate of first reproduction (HR = 1.072, p = 0.005), and maternal loss more than doubled the rate of first birth (HR = 2.411, p = 0.020). Critically, this acceleration was specific to psychosocial stressors; no ecological or energetic variable predicted reproductive timing.

Multiple linear regression models examining life history outcomes among parous females yielded sparse but informative results. Only models in the first-five-year window reached overall significance. Cumulative absolute rainfall deviation predicted both shorter reproductive lifespan

(β = −9.04, FDR p = 0.021) and lower offspring survival (β = −0.644, FDR p = 0.016). Group-size in the first-year also predicted lower offspring survival (β = −0.00218, FDR p = 0.009), but with an extraordinarily small effect size that makes it difficult to interpret. The prenatal window yielded no significant predictors after false discovery rate correction in any model.

Taken together, the Chapter 1 results reveal that developmental constraint and predictive calibration are not competing explanations but concurrent processes operating at different timescales and through different adversity types. Ecological adversity constrains, kills early and – among survivors – shortens reproductive lifespan and reduces offspring survival. Psychosocial adversity also kills early, while the six-year AFB sensitivity analysis provides suggestive evidence that it may additionally influence reproductive timing among survivors, consistent with PAR predictions. The contrast between from-birth and delayed-entry mortality models further demonstrates that viability selection can obscure developmental effects when analyses condition on survival to later life.

Chapter 2: Boldness, Aggression, and the Behavioral Syndrome

Chapter 2 characterized individual differences in boldness and aggression as behavioral phenotypes using directly observed measures collected across the lifespan. The chapter evaluated whether these traits constitute repeatable individual differences, whether they covary at the population level to form a boldness–aggression syndrome, and whether the population-level analysis conceals individual heterogeneity in how the two traits are coupled. It also evaluated the convergent validity of observer-assigned personality ratings with the directly measured behavioral traits.

The results challenged the assumptions underlying the behavioral syndrome literature in

ways that are theoretically informative rather than merely disappointing. Boldness, operationalized as approach distance during standardized encounters with snake models, showed very low repeatability ($R \sim 0.02–0.09$), substantially below the meta-analytic mean of approximately 0.37 for behavioral traits in free-living animals (Bell et al., 2009) and below the range typically cited as evidence for a stable personality trait. Aggression was target-specific and context-dependent rather than trait-like in the manner the syndrome literature assumes: rates directed at different social targets were not reliably correlated, and a single individual-level aggression parameter did not capture meaningful consistency across interaction types. No population-level boldness–aggression syndrome emerged from the data. However, these are not simply findings of low statistical power: the individual-level heterogeneity analysis showed that some individuals do show positive coupling between boldness and aggression, others show negative coupling, and the population correlation is an uninformative average across individuals whose behavioral architectures genuinely differ.

The finding that bolder individuals show greater behavioral flexibility, defined as smaller variance across repeated snake encounters rather than larger variance, directly inverts the prediction of proactive–reactive coping styles theory, which associates boldness with behavioral rigidity and routine formation (Koolhaas et al., 1999, 2010). In capuchins, where foraging success, coalition maintenance, and competitive success all demand behavioral inventiveness across changing social and ecological contexts, boldness toward novel or risky stimuli may reflect generalized engagement with the environment rather than the inflexible proactive strategy that the rodent-derived framework describes. This inversion is a substantive positive finding about the structure of behavioral variation in a cognitively complex primate, not an uninformative null result.

Chapter 3: Estimating GC Rection Norms

Chapter 3 addressed a fundamental methodological challenge in field endocrinology: how to extract biologically meaningful tonic and reactive HPA axis phenotypes from fGCM data collected opportunistically across many years of naturalistic observation. The chapter developed a rolling-window methodology that leverages the species-specific fGCM peak excretion window to separate samples reflecting tonic HPA axis activity from those reflecting acute HPA axis activation in response to intergroup encounters. It then characterized the repeatability of each HPA axis dimension and tested whether they are empirically independent.

Two findings from this chapter bear directly on the theoretical framework and on the analyses in Chapter 4. First, tonic HPA axis activity is a repeatable individual phenotype, with an intraclass correlation coefficient of 0.295, well within the range of 0.20–0.45 documented in meta-analyses of baseline glucocorticoid repeatability in free-living vertebrates (Schoenemann & Bonier, 2018; Taff et al., 2018). This establishes that individual differences in baseline glucocorticoid activity are real, stable, and large enough to serve as a trait-level predictor in downstream analyses. Second, and critically for the Adaptive Calibration Model's four-pattern taxonomy, tonic and reactive HPA axis activity were empirically uncorrelated ($r \sim -0.07$). The near-zero correlation between the two dimensions demonstrates little covariance between these two phenotypes within this analysis, consistent with the possibility that they represent genuinely independent axes of HPA axis regulation, rather than redundant expressions of a single underlying glucocorticoid trait. An individual's typical baseline glucocorticoid output may carry little information about how that same individual will respond to an acute stressor.

The rolling-window methodology developed here has implications beyond the capuchin system. It represents a principled approach to extracting temporally distinct HPA axis signals from

naturalistic field data that avoids both the imprecision of treating all fGC values as equivalent and the logistical infeasibility of experimental manipulation in wild populations. The methodological validation work, including simulation analyses characterizing how sampling depth affects the reliability of individual slope estimates for acute reactivity, provides an honest accounting of what this approach can resolve with current sampling density, and points toward the design features that future field studies would need to improve HPA axis phenotyping in wild primates.

Chapter 4: The Full Causal Chain in Males

Chapter 4 integrated the foregoing components to test the full hypothesized causal chain in males: whether ELA programs adult HPA axis function, whether HPA axis phenotypes predict behavioral traits, whether those traits predict male life history transitions, and whether any mediated pathway from ELA through physiology and behavior to fitness is supported. The results speak to each link in the chain in turn.

The ELA → HPA axis link was largely absent. Of six first-year adversity indicators examined, only elevated group mortality rate predicted lower tonic glucocorticoid activity in adulthood ($b = -0.644$, FDR-corrected $p = 0.035$). No adversity indicator robustly predicted HPA axis reactivity after correction for multiple comparisons. Two marginal trends, rainfall deviation on reactivity and alpha-male takeover exposure on reactivity, were suggestive in sensitivity analyses but did not survive multiple comparison correction. Maternal loss, which is the adversity type most consistently associated with HPA axis programming in the experimental primate literature (Maestripieri et al., 2006; Parker & Maestripieri, 2011), could not be evaluated in the HPA axis analysis because prevalence in the first-year analytic sample was too low (owing to the fact that individuals who experience maternal loss in the first year of life face much higher odds

of dying prior to adulthood). The overall footprint of early adversity on adult HPA axis function was therefore narrow and modest: one robust signal in one direction (higher early mortality → lower tonic activity), and no robust effects on acute reactivity.

The HPA axis → behavior link was present for one trait and absent for the other. Higher tonic glucocorticoid activity was associated with greater boldness at the level of stable individual differences (among-individual $r = -0.36$, 95% CI [−0.57, −0.08], $p = 0.010$), a genuine physiology–behavior covariance. But the direction of this association inverted the proactive–reactive prediction: higher tonic cortisol corresponded to bolder, not more cautious behavior. Neither tonic nor reactive HPA axis activity predicted proactive or reactive aggression (all $|r| < 0.14$, all $p > .43$). The physiological substrate that underlies boldness variation in this population does not extend to the social-competitive behavioral domain, a finding that reinforces the modularity of the behavioral phenotype and challenges the assumption of a unified physiological personality axis.

The ELA → behavior pathway (bypassing HPA axis) produced the chapter's clearest positive finding. Maternal loss during the first five years of life predicted both greater adult boldness ($b = -0.10$, $p = 0.032$) and reduced within-individual variability in boldness responses ($b_\sigma = -0.118$, $p = 0.022$). Maternally bereaved individuals were not only bolder on average, but also more consistent: their risk-taking behavior was canalized into a narrow, reliably bold phenotype less likely to flex in response to contextual variation across snake encounters. This interpretation aligns with developmental perspectives on animal personality, which emphasize that early experience can organize stable individual differences in behavior, and with empirical work showing that stressful developmental conditions can reduce within-individual behavioral variation and increase behavioral consistency (Royauté & Dochtermann, 2017; Stamps &

Groothuis, 2010).

The behavior → life history link was present for proactive aggression and absent for all other predictors. Proactive aggression was the primary behavioral predictor of alpha attainment, with males higher in proactive aggression showing more than twice the hazard of attaining alpha status at any given age (HR = 2.22, 95% CI [1.14, 4.35], p = 0.020). Reactive aggression, boldness, and tonic HPA axis activity were null predictors of attainment timing. Among ELA predictors, alpha-male instability during the first five years showed a positive but marginal association with earlier alpha attainment (HR = 1.55, 95% CI [1.08, 2.24]), suggesting a possible developmental effect on competitive trajectory, though the mechanism is unresolved. Maternal loss was associated with earlier natal emigration (HR = 2.15, 95% CI [1.34, 3.46], p = 0.002), while ecological harshness during early development was associated with delayed emigration (HR = 0.58, 95% CI [0.40, 0.83], p = 0.003), a divergence between adversity types that aligns with findings from Chapter 1 and further disconfirms any uniform "adversity → accelerated life history" prediction.

The full mediated pathway from ELA through HPA axis to behavior to life history was not supported. The two conditions required for HPA axis-mediated effects, that ELA predicts HPA axis function, and that HPA axis function predicts life history outcomes, were both largely unsatisfied. HPA axis phenotypes predicted no life history outcome across emigration timing, alpha attainment, or alpha tenure duration (all hazard ratios within 0.90–1.36, all confidence intervals spanning 1.0). The HPA axis, as characterized by fecal metabolite measurement, is not functioning as the physiological conduit through which early adversity reaches male fitness outcomes in this population.

II. Synthesis

The failure of the ELA → HPA axis → behavior → life history chain to hold is not attributable to any single methodological limitation. Instead, it reflects a substantive mismatch between the assumptions that underlie the standard framework and the features of the capuchin system. Three assumptions deserve particular scrutiny.

The first is the assumption of predictive cue reliability. Informational models of adaptive developmental plasticity, including PAR frameworks and the ACM, require that early environmental conditions carry reliable statistical information about the conditions an organism will encounter as an adult, information that natural selection can exploit by calibrating the developing physiological system to the anticipated future environment (Del Giudice et al., 2011; Gluckman et al., 2005; Nettle & Bateson, 2015). In white-faced capuchins, this requirement faces a serious obstacle: lifespan exceeds thirty years, the juvenile period spans approximately six to seven years, and the social and ecological environments are continuously fluctuating, driven by ENSO-linked rainfall variability, coalitionary realignments, group fissions, and alpha male turnover events (Perry, 2012).

Under these conditions, the correlation between the social and ecological environment of infancy and the environment of reproduction is likely to be very low. Theoretical work on the evolution of cue-based plasticity has shown that selection for predictive calibration weakens rapidly as the lag between developmental cue and expressed phenotype increases and as adult environmental conditions become increasingly unpredictable (Fawcett & Frankenhuis, 2015; Frankenhuis et al., 2019); both are conditions that apply to capuchin life history. Thus, the Chapter 1 finding that psychosocial adversity accelerates reproductive timing among surviving females is unlikely to reflect a straightforward predictive adaptive response in which early social conditions accurately forecast the adult reproductive environment. Instead, psychosocial

adversity may operate through an internal state-dependent pathway. Early maternal loss, social instability, or other socially embedded stressors may alter the developing stress axis and its interaction with the reproductive axis, shifting the timing of adrenarche, pubertal maturation, or reproductive readiness. Early reproduction is not observed because females have "predicted" the future environment, but because early adversity has reorganized physiological development in ways that accelerate life-history transitions. The result may resemble a PAR at the phenotypic level, while arising through internal calibration of somatic, neuroendocrine, and reproductive systems rather than through reliable external forecasting.

The second is the assumption of behavioral syndrome coherence. Much of the theoretical framework linking HPA axis function to life history outcomes operates through a latent "fast–slow" axis of behavioral strategy, a proactive coping style presumed to cluster boldness, high aggression, low glucocorticoid output, and behavioral rigidity into a coherent syndrome that maps onto a fast life history pace (Koolhaas et al., 1999, 2010; Réale et al., 2007, 2010). The empirical landscape in capuchins does not support this structure. Chapter 2 found no population-level boldness–aggression syndrome, with individual heterogeneity in the coupling of the two traits as large as the average coupling itself. Chapter 4 found that the two behavioral traits were predicted by different antecedents - maternal loss shapes boldness and its consistency; proactive aggression is the relevant predictor of competitive success - and neither trait shared a physiological basis with the other in the form of a common HPA axis substrate. Rather than leading to an integrated response, adversity leaves domain-specific rather than syndrome-wide marks on behavior. Physiological covariation with behavior is trait-specific rather than reflecting a general personality axis. This modularity is itself theoretically interpretable in the context of a cognitively complex species whose fitness depends on behavioral flexibility across multiple

independent domains simultaneously (Coppens et al., 2010; Fragaszy et al., 2004; Perry et al., 2003; Sih & Del Giudice, 2012).

The third is the assumption that the HPA axis is the primary physiological integrator of developmental experience. This assumption follows naturally from the experimental literature, in which early adversity paradigms reliably program glucocorticoid regulation through well-characterized molecular mechanisms (Liu et al., 1997; Meaney, 2001; Weaver et al., 2004). In that literature, the HPA axis is the natural target because the experimental manipulations directly activate the stress response system. The adversities that occur naturally in wild capuchin populations are structurally different. Prenatal takeover exposure is a socially mediated threat that may reach the fetus through glucocorticoid transfer but whose effects on mortality are likely to travel through the immediate threat of infanticide. Maternal loss operates through attachment disruption, resource loss, reduced immune protection, and social exposure, and its primary detectable downstream effect in Chapter 4 is on the canalization of boldness, not on the glucocorticoid axis. Ecological harshness during early life affects energetic state, competitive capacity, and the timing of high-risk developmental transitions, but these effects appear to operate through energetic and somatic channels that do not register as persistent differences in fGCM concentration. None of this means the HPA axis is irrelevant, but rather that treating it as the primary or exclusive physiological integrator of developmental experience may be an artifact of the experimental framework rather than a feature of the biology.

III. Persistent Developmental Effects

The findings that replicate across analyses, that (a) early-life adversity is lethal and viability selection removes the most affected, (b) maternal loss canalizes boldness in males, (c) proactive

aggression predicts alpha attainment, and (d) tonic and reactive HPA axis are independent, are specific, theoretically grounded positive findings that collectively advance understanding of how developmental experience, physiology, and behavior each independently shape fitness outcomes in a long-lived social primate. The six-year AFB sensitivity analysis additionally raises the possibility that psychosocial adversity influences reproductive timing among surviving females.

One of the most consequential findings for interpreting developmental programming in this dissertation is the evidence for strong viability selection. In from-birth survival models, prenatal alpha male replacement quadrupled female mortality risk (HR = 4.34), and maternal loss during the first year produced a comparable elevation (HR = 4.43). Although the lethality of early-life adversity itself is well established, its magnitude here has important implications for attempts to detect its downstream consequences. Specifically, individuals most strongly affected by adversity may die before adult physiological or behavioral phenotypes can be measured. Consistent with this possibility, these mortality associations vanished when conditioning on survival to age five. The most adversity-exposed individuals die during the juvenile period, and any analysis that begins at adulthood inherits a sample from which the mortality signal has already been purged. This finding has implications well beyond the capuchin system. Human epidemiological cohorts, which necessarily condition on survival to recruitment, may systematically underestimate the fitness consequences of early-life adversity for the same reason.

The suggestion that psychosocial adversity accelerates reproductive timing among surviving females adds a potentially important dimension to the Chapter 1 story and creates an interesting point of comparison to the Chapter 4 male findings. In the six-year AFB sensitivity analysis, alpha male replacements, and maternal loss during the first five years significantly predicted earlier first birth (HR = 1.072 per replacement, $p = 0.005$; HR = 2.411 for maternal

loss, $p = 0.020$), while no ecological or energetic variable reached significance. This contrast is theoretically important because energetic stress and psychosocial stress should not be expected to affect reproductive timing through the same pathway. Energetic stress can directly constrain reproductive potential by limiting the metabolic resources available for growth, ovarian function, and successful reproduction, consistent with reproductive ecology models in which female reproductive function is highly sensitive to energy availability (Ellison, 2003; Jasienska & Ellison, 2004). Psychosocial stressors, by contrast, need not impose the same immediate energetic ceiling. The association between psychosocial adversity and earlier reproductive timing in the six-year sensitivity analysis could arise through pathways distinct from energetic constraint. One possibility is that socially embedded stressors alter aspects of somatic or neuroendocrine development that influence the timing of reproductive maturation. Importantly, however, the present analyses do not identify the mechanism underlying this association. Thus, although the observed pattern resembles the accelerated reproductive timing predicted by PAR models, it does not by itself demonstrate that females use early social conditions to forecast the future environment. Distinguishing predictive calibration from state-dependent developmental effects will require direct measures of physiological processes linking early social adversity to reproductive maturation.

The canalization of boldness following maternal loss remains one of the clearest behavioral findings of the dissertation. Developmental approaches to animal personality emphasize that early experience can shape stable individual differences in behavior, including both average behavioral tendencies and the degree to which individuals vary across contexts (Stamps & Groothuis, 2010). On this view, early adversity could plausibly produce either heightened behavioral flexibility, if individuals become more responsive to environmental variation, or canalization, if adversity

narrows the range of later behavioral responses. The capuchin data support the latter interpretation: males who lost their mothers early in life were not only bolder on average, but also less variable in their boldness responses. This pattern suggests that maternal loss does not simply shift males toward greater risk-taking; it also constrains the behavioral repertoire within the boldness domain. Rather than indicating adaptive forecasting of future conditions, the result is better interpreted as a domain-specific developmental constraint, in which disruption of the primary social relationship during a sensitive period produces a bold phenotype that is elevated but behaviorally rigid.

The finding that proactive aggression predicts alpha attainment is the clearest demonstration in this dissertation that behavioral phenotypes matter for fitness in male capuchins and through specific, domain-relevant channels rather than through a general fast-life personality axis. An individual's propensity for unprovoked, instrumentally motivated aggression prior to alpha status is what best predicts whether and when he achieves that status (HR = 2.22), with reactive aggression, boldness, and HPA axis phenotypes all null. Competitive success is a consequence of a specific behavioral strategy, not a general personality type.

Finally, the demonstration that tonic and reactive HPA axis activity are empirically independent in this population carries methodological and theoretical implications that extend well beyond this dataset. Methodologically, it validates the rolling-window approach as a tool for extracting dimensionally distinct HPA axis phenotypes from naturalistic field data. Theoretically, it provides direct field support for the ACM's core claim that tonic and reactive HPA axis activity can be configured independently while simultaneously demonstrating that this independence does not by itself generate the coordinated behavioral profiles the ACM predicts.

IV. Sex-Specific Pathways from Adversity to Fitness

A central organizing principle of this dissertation is that the pathway from early-life adversity to fitness is sex-specific because the proximate determinants of fitness differ between the sexes in ways that redirect the consequences of adversity into different phenotypic domains. The findings across all four chapters, when read together, support and refine this principle.

Female reproductive output is constrained primarily by the female's own physiology: her energetic state, somatic condition, reproductive lifespan, and the hormonal regulation of ovarian function (Ellison, 2003). The Chapter 1 analyses reveal that early adversity affects female fitness through two concurrent pathways. The first is through increased mortality risk in the immediate aftermath - both psychosocial and ecological adversity kill during the juvenile period. The second, visible only among females who survive this mortality filter, is a developmental programming pathway: psychosocial adversity, specifically alpha male replacements, and maternal loss, predicts earlier reproductive timing, though this effect is only observed in the 6-year AFB sensitivity analysis. This acceleration should not be interpreted as evidence that social adversity provides reliable information about future reproductive conditions, nor does the present analysis establish that earlier reproduction represents an adaptive response to adversity. Instead, the association may reflect state-dependent developmental effects of psychosocial stress on reproductive maturation. Such effects could arise through changes in somatic or neuroendocrine development, but these potential mechanisms were not measured here. Thus, the observed pattern is consistent with accelerated reproductive timing following psychosocial adversity without distinguishing whether that acceleration reflects adaptive calibration, developmental constraint, or some other consequence of early life social stress.

In males, fitness is primarily constrained by access to mating opportunities, and in this species those opportunities are concentrated at the alpha position (Perry et al., 2012). Because

reproductive success is so highly skewed, the relevant male life history transitions are not simply energetic or somatic thresholds for reproduction but competitive transitions: dispersal, rise to alpha male status, and fall from alpha male status. Pathways from early adversity to male fitness may therefore be especially likely to involve phenotypes relevant to social competition. Consistent with this expectation, the clearest behavioral life history finding (proactive aggression predicts alpha attainment) involves a social-competitive behavioral trait rather than a physiological marker, while the strongest direct adversity effect on a male life history transition (maternal loss accelerating emigration, ecological harshness delaying it) operates through competitive social readiness rather than reproductive physiology. These findings do not exclude physiological mediation of male competitive trajectories; rather, they indicate that the HPA axis dimensions measured here did not account for these associations. Other physiological systems involved in reproductive and competitive function (notably, androgens) were not evaluated.

The Chapter 1 results also raise an intriguing potential parallel in the sex-specific pathways. In both sexes, psychosocial adversity (particularly maternal loss) accelerates a key life history transition: the onset of reproduction in females (HR = 2.41 for maternal loss predicting earlier AFB) and the departure from the natal group in males (HR = 2.15 for maternal loss predicting earlier emigration). These are functionally analogous transitions, both represent the point at which an individual begins investing in the next phase of its reproductive career (though male capuchins may begin reproducing within their natal group, this is much less likely). Because the female association was not detected in the primary AFB models, and the mechanisms underlying either association were not directly tested, this pattern should be treated as a hypothesis for further investigation rather than evidence for a shared developmental mechanism. Future work could test

whether maternal loss influences the timing of life-history transitions across sexes and, if so, whether common social, physiological, or developmental processes account for that pattern.

Together, the two sex-specific pathways reveal that "adversity effects on fitness" is not a single phenomenon but a family of related phenomena whose proximate nature depends on what determines fitness in a given sex-class. The search for universal mechanisms linking ELA to life history outcomes may be systematically misleading when the actual fitness pathways are sex-specific and involve partly non-overlapping phenotypic mediators. This dissertation demonstrates that decomposing the ELA-to-fitness pathway by sex and characterizing the proximate determinants of fitness in each sex-class before formulating predictions about mediating mechanisms, is a necessary step for making predictions that are actually testable in wild populations.

V. Limitations

Several limitations of the present analyses constrain the inferences that can be drawn and motivate specific priorities for future work.

The most consequential limitation is the inability to evaluate the effect of maternal loss on HPA axis function - the single adversity type for which the experimental literature most consistently documents developmental programming of the glucocorticoid axis (Maestripieri et al., 2006; Parker & Maestripieri, 2011; Sanchez, 2006). In the Chapter 4 HPA axis analyses, the first-year maternal loss analytic sample contained too few exposed individuals for the model to converge, reflecting the fact that surviving maternal loss in the first year of life is uncommon in this population and that the subset of individuals with sufficient adult fGCM data to enter HPA axis analyses was smaller than the full population. The most significant gap in the ELA → HPA

axis analysis is therefore not a test that failed but a test that could not be conducted. Future work with expanded HPA axis sampling, either by extending the measurement period or by supplementing opportunistic sampling with targeted collection around known adversity exposures, could fill this gap.

A second limitation concerns the measurement of acute HPA axis reactivity. The rolling-window methodology developed in Chapter 3 identified a biologically plausible post-intergroup-encounter signal in the early post-stressor window, but per-individual sampling in that window was sparse (median of approximately two post-encounter samples per individual in the rise window), and simulation analyses indicated a correlation of only 0.46 between true and estimated individual reactive slopes at this sampling depth. Null findings involving the reactive HPA axis phenotype should therefore be interpreted as inconclusive rather than as positive evidence that individuals do not differ in acute responsiveness. The predictions of the ACM regarding coordinated tonic–reactive profiles could not be fully evaluated with the current dataset because one of the two required dimensions could not be characterized with adequate reliability at the individual level. Targeted collection of fecal samples in the one-to-four-hour window following identifiable stressors is the most direct way to improve this.

Third, the behavioral analyses in Chapter 4 rely on individual-level behavioral predictors derived as best linear unbiased predictors (BLUPs) from random effects models. When used as predictors in downstream survival analyses, BLUP-derived regressors are subject to generated-regressor attenuation bias proportional to the square root of the trait's repeatability. Given the low repeatability of boldness documented in Chapter 2 ($R \sim 0.02–0.09$) and the moderate repeatability of tonic HPA axis (ICC ~ 0.295), the two-step BLUP approach deflates effect size estimates for paths where a BLUP is the predictor. The bivariate model used for the tonic HPA axis–boldness

correlation avoids this problem by estimating the covariance parameter directly rather than conditioning on a point estimate, but the Cox models predicting life history transitions from behavioral and physiological BLUPs do not. The null findings for HPA axis predicting life history outcomes should therefore be read as conservative lower bounds on the true effect size rather than as strong evidence against an HPA axis–life history association.

Fourth, sample sizes for the rarest life history outcomes are small. The alpha attainment analysis comprised 56 males with 13 attainment events, and the alpha tenure analysis comprised 17 males with 11 tenure-ending events. Effect size estimates for behavioral predictors of these outcomes, while statistically significant for proactive aggression predicting attainment, carry wide confidence intervals and should be interpreted as provisional. The directional pattern of findings is internally consistent across the chapter, but replication with a larger sample of males reaching alpha status, requiring at least an additional decade or more of ongoing data collection, would considerably strengthen the inference.

Fifth, the dissertation examined the HPA axis as the candidate physiological mediator of developmental programming, to the exclusion of other potentially relevant physiological systems. The hypothalamic–pituitary–gonadal axis, sympathetic–adrenal–medullary function, immune markers, and metabolic indicators are all plausible alternative or complementary physiological intermediates through which early adversity might shape adult phenotype and fitness outcomes. The failure of HPA axis-mediated effects to materialize in the capuchin data does not rule out the possibility that other systems carry the developmental signal; it simply establishes that the HPA axis, as characterized by fGCM metabolites, is not the primary conduit in this population and with this set of adversity measures.

VI. Future Directions

The findings and limitations of this dissertation point toward several specific priorities for advancing the study of developmental programming in wild primates.

The most pressing methodological priority is targeted post-stressor fecal sampling. The rolling-window approach developed in Chapter 3 demonstrated that the conceptual separation of tonic and reactive HPA axis is achievable with field data, but the current opportunistic sampling protocol cannot reliably characterize individual differences in acute reactivity. Research teams at long-term primate field sites with identified stressor events would benefit enormously from a sampling protocol that prioritizes collection in the early post-stressor window. Even a modest number of deliberately timed samples per individual would substantially improve the reliability of reactive phenotype estimates and enable a genuine test of the ACM's tonic–reactive profile predictions.

A second priority is the extension of developmental analyses to additional physiological systems. Gonadal hormones are particularly promising: testosterone is a direct regulator of competitive behavior and dominance acquisition in male primates, and the developmental programming of the HPG axis by early social adversity is well documented in captive studies (Sapolsky et al., 2000). Whether naturally occurring adversity in this population programs adult testosterone levels or cortisol–testosterone dynamics, and whether gonadal hormones mediate the relationship between proactive aggression and alpha attainment more robustly than glucocorticoids do, are open questions that could be addressed with existing fecal hormone metabolite techniques applied to the same archived samples that yielded the glucocorticoid data.

A third priority is the longitudinal tracking of boldness across developmental stages. The canalization finding from Chapter 4 raises a question the current data cannot answer: when does the effect of maternal loss on boldness consistency emerge? If canalization is a gradual process that crystallizes across the juvenile period as the individual assembles a coping strategy without maternal support, we would expect boldness variance to decline over time in bereaved individuals but not in non-bereaved controls. If instead canalization reflects an acute developmental perturbation at the time of loss, variance reduction should be detectable relatively soon after the loss event. Distinguishing these temporal profiles would clarify whether canalization reflects constraint (immediate narrowing of the developmental trajectory) or state-dependent adaptation (gradual commitment to a high-risk strategy under conditions of social vulnerability).

A fourth priority is the evaluation of coalition structure as a mediating variable in the ELA → life history pathway. The finding that alpha male instability during early development is associated with earlier alpha attainment (HR = 1.55) raises the possibility that exposure to frequent competitive disruption during development shapes not just individual behavioral traits but the quality and structure of the coalitionary networks that determine male competitive success (Gros-Louis et al., 2003; Kajokaite et al., 2019; Perry, 1998). Individuals who grow up observing frequent coalitionary transitions may acquire specific social knowledge, about ally reliability, competitive tactics, or rank dynamics, which shapes their competitive trajectory more directly than any hormonal or personality phenotype. Network-analytic approaches to the male social data, examined alongside developmental history, could test this hypothesis.

VII. Contribution and Closing Statement

This dissertation makes several contributions to the scientific literature that extend beyond the specific findings for white-faced capuchins at Lomas Barbudal. At the most general level, it provides the most comprehensive test to date of the full ELA → HPA axis → behavior → life history causal chain in a wild, non-human primate population with individually known developmental histories. The test is comprehensive in the sense that it evaluates all four links simultaneously, uses prospective rather than retrospective adversity records, and covers a range of adversity types, physiological dimensions, behavioral traits, and life history outcomes in a single dataset. The coherent pattern of what fails and what holds is more interpretable than the results of any single-link study and provides the kind of evidence that theory revision requires.

A central theoretical contribution of the Chapter 1 analyses is the demonstration that developmental constraint and predictive adaptive calibration are not competing frameworks but concurrent processes operating at different timescales and through different adversity types in the same population. Ecological adversity kills early and constrains life history among survivors. Psychosocial adversity also kills early but additionally calibrates reproductive timing among survivors. The delayed-entry null results that would be the standard finding in most developmental origins studies are revealed to be an artifact of viability selection, not evidence of absent programming. This dual-process framework resolves an apparent contradiction in the broader literature, why some studies find constraint while others find calibration, and suggests that the type of adversity, not just its intensity, determines which process dominates.

The methodological contribution of the rolling-window HPA axis phenotyping approach is likely to have utility beyond this system. Long-term primate field studies with archived fecal hormone data are increasingly common, and the challenge of separating tonic from reactive HPA axis signals in opportunistically collected samples is shared across virtually all of them. A

validated, simulation-tested methodology for making this separation with explicit characterization of the reliability limits imposed by typical sampling depth fills a gap in field endocrinology methodology that has limited the comparability of findings across sites and species.

The theoretical contribution extends beyond the dual-process framework to a specific, empirically grounded critique of the conditions under which informational calibration models can be expected to hold. The data demonstrate that the particular mechanism most prominently featured in current frameworks is not the mechanism at work in this population. The failure is theoretically informative because it maps onto the conditions under which that mechanism *should fail*: long lifespan, fluctuating environments, high cue-to-expression lag, and a fitness landscape dominated by coalitionary social dynamics (Fawcett & Frankenhuis, 2015; Frankenhuis et al., 2019; Sih & Del Giudice, 2012). Yet the partial success of PAR predictions for psychosocial adversity in females suggests that the framework is not entirely wrong; rather, it is too coarse in its treatment of adversity types and too optimistic about the generality of HPA axis mediation.

The specific character of what does hold, that (a) early adversity is lethal and viability selection purges the most affected; (b) psychosocial adversity accelerates reproduction in surviving females and emigration in males; (c) maternal loss canalizes boldness; and (d) proactive aggression builds alpha careers, suggests that the more productive theoretical direction is not to search for a single integrating physiological mechanism that translates adversity into life history strategy, but to characterize how specific forms of adversity leave specific marks on specific phenotypic domains relevant to fitness in each sex. The sex-specific pathway logic developed in this dissertation, strengthened by the parallel finding that psychosocial adversity accelerates key life transitions in both sexes, is a step toward a more ecologically grounded framework for developmental programming in primates. Extending that logic to other species with different social

structures and fitness landscapes should be a priority for comparative work. White-faced capuchins at Lomas Barbudal, with their extraordinary longitudinal record, their coalitionary complexity, their slow development, and their naturally occurring variation in every adversity type examined here, are an ideal system for continuing that work and the three decades of data that make this dissertation possible are a foundation, not a ceiling, for what can be learned.

References

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

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 B: Biological Sciences*, *365*(1560), 4021–4028. https://doi.org/10.1098/rstb.2010.0217

Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The Adaptive Calibration Model of stress responsivity. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1562–1592. https://doi.org/10.1016/j.neubiorev.2010.11.007

Ellison, P. T. (2003). Energetics and reproductive effort. *American Journal of Human Biology*, *15*(3), 342–351. https://doi.org/10.1002/ajhb.10152

Fawcett, T. W., & Frankenhuis, W. E. (2015). Adaptive explanations for sensitive windows in development. *Frontiers in Zoology*, *12*(Suppl 1).

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The Complete Capuchin: The Biology of the Genus Cebus*. Cambridge University Press.

Frankenhuis, W. E., Nettle, D., & Dall, S. R. X. (2019). A case for environmental statistics of early-life effects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1770), 20180110. https://doi.org/10.1098/rstb.2018.0110

Gluckman, P. D., Hanson, M. A., & Spencer, H. G. (2005). Predictive adaptive responses and human evolution. *Trends in Ecology & Evolution*, *20*(10), 527–533. https://doi.org/10.1016/j.tree.2005.08.001

Gros-Louis, J., Perry, S., & Manson, J. H. (2003). Violent coalitionary attacks and intraspecific killing in wild white-faced capuchin monkeys (Cebus capucinus). *Primates*, *44*, 341–346.

Jasienska, G., & Ellison, P. T. (2004). Energetic factors and seasonal changes in ovarian function in women from rural Poland. *American Journal of Human Biology*, *16*(5), 563–580. https://doi.org/10.1002/ajhb.20063

Kajokaite, K., Whalen, A., Panchanathan, K., & Perry, S. (2019). White-faced capuchin monkeys use both rank and relationship quality to recruit allies. *Animal Behaviour*, *154*, 161–169. https://doi.org/10.1016/j.anbehav.2019.06.008

Koolhaas, J. M., de Boer, S. F., Coppens, C. M., & Buwalda, B. (2010). Neuroendocrinology of coping styles: Towards understanding the biology of individual variation. *Frontiers in Neuroendocrinology*, *31*, 307–321.

Koolhaas, J. M., Korte, S. M., de Boer, S. F., Van Der Vegt, B. J., Van Reenen, C. G., Hopster, H., De Jong, I. C., Ruis, M. A. W., & Blockhuis, H. J. (1999). Coping styles in animals: Current status in behavior and stress-physiology. *Neuroscience & Biobehavioral Reviews*, *23*, 925–935.

Liu, D., Diorio, J., Tannenbaum, B., Caldji, C., Francis, D., Freedman, A., Sharma, S., Pearson, D., Plotsky, P. M., & Meaney, M. J. (1997). Maternal Care, Hippocampal Glucocorticoid Receptors, and Hypothalamic-Pituitary-Adrenal Responses to Stress. *Science*, *277*(5332), 1659–1662. https://doi.org/10.1126/science.277.5332.1659

Maestripieri, D., Higley, J. D., Lindell, S. G., Newman, T. K., McCormack, K. M., & Sanchez, M. M. (2006). Early maternal rejection affects the development of monoaminergic systems and adult abusive parenting in rhesus macaques (Macaca mulatta). *Behavioral Neuroscience*, *120*(5), 1017–1024. https://doi.org/10.1037/0735-7044.120.5.1017

Meaney, M. J. (2001). Maternal Care, Gene Expression, and the Transmission of Individual Differences in Stress Reactivity Across Generations. *Annual Review of Neuroscience*, *24*(1), 1161–1192. https://doi.org/10.1146/annurev.neuro.24.1.1161

Nettle, D., & Bateson, M. (2015). Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151005. https://doi.org/10.1098/rspb.2015.1005

Nettle, D., Frankenhuis, W. E., & Rickard, I. J. (2013). The evolution of predictive adaptive responses in human life history. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1766), 20131343. https://doi.org/10.1098/rspb.2013.1343

Parker, K. J., & Maestripieri, D. (2011). Identifying key features of early stressful experiences that produce stress vulnerability and resilience in primates. *Neuroscience & Biobehavioral Reviews, Resilience and Adaptive Aspects of Stress in Neurobehavioural Development*, *35*(7), 1466–1483. https://doi.org/10.1016/j.neubiorev.2010.09.003

Perry, S. (1998). Male-Male Social Relationships in Wild White-Faced Capuchins, Cebus capucinus. *Behaviour*, *135*(2), 139–172.

Perry, S., Baker, M., Fedigan, L., Gros-Louis, J., Jack, K., MacKinnon, K. C., Manson, J. H., Panger, M., Pyle, K., & Rose, L. (2003). Social Conventions in Wild White-faced Capuchin Monkeys: Evidence for Traditions in a Neotropical Primate. *Current Anthropology*, *44*(2), 241–268. https://doi.org/10.1086/345825

Perry, S., Godoy, I., & Lammers, W. (2012). The Lomas Barbudal Monkey Project: Two Decades of Research on Cebus capucinus. In P. M. Kappeler & D. P. Watts (Eds.), *Long-Term Field Studies of Primates* (pp. 141–163). Springer. https://doi.org/10.1007/978-3-642-22514-7_7

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P.-O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4051–4063. https://doi.org/10.1098/rstb.2010.0208

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. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Royauté, R., & Dochtermann, N. A. (2017). When the mean no longer matters: Developmental diet affects behavioral variation but not population averages in the house cricket (*Acheta domesticus*). *Behavioral Ecology*, *28*(1), 337–345. https://doi.org/10.1093/beheco/arw164

Sanchez, M. M. (2006). The impact of early adverse care on HPA axis development: Nonhuman primate models. *Hormones and Behavior, Translational Topics in Behavioral Neuroendocrinology*, *50*(4), 623–631. https://doi.org/10.1016/j.yhbeh.2006.06.012

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How Do Glucocorticoids Influence Stress Responses? Integrating Permissive, Suppressive, Stimulatory, and Preparative Actions*. *Endocrine Reviews*, *21*(1), 55–89. https://doi.org/10.1210/edrv.21.1.0389

Schoenemann, K. L., & Bonier, F. (2018). Repeatability of glucocorticoid hormones in vertebrates: A meta-analysis. *PeerJ*, *6*, e4398. https://doi.org/10.7717/peerj.4398

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.

Stamps, J., & Groothuis, T. G. G. (2010). The development of animal personality: Relevance, concepts and perspectives. *Biological Reviews*, *85*(2), 301–325. https://doi.org/10.1111/j.1469-185X.2009.00103.x

Taff, C. C., Schoenle, L. A., & Vitousek, M. N. (2018). The repeatability of glucocorticoids: A review and meta-analysis. *General and Comparative Endocrinology*, *260*, 136–145. https://doi.org/10.1016/j.ygcen.2018.01.011

Weaver, I. C. g., Diorio, J., Seckl, J. R., Szyf, M., & Meaney, M. J. (2004). Early Environmental Regulation of Hippocampal Glucocorticoid Receptor Gene Expression: Characterization of Intracellular Mediators and Potential Genomic Target Sites. *Annals of the New York Academy of Sciences*, *1024*(1), 182–212. https://doi.org/10.1196/annals.1321.099