2019

Measuring Behavioral Phenotypes From Complex Social Interactions In Non-Human Primates

Seth Madlon-Kay
University of Pennsylvania

Measuring Behavioral Phenotypes From Complex Social Interactions In Non-Human Primates

Abstract

Humans are innately social animals, and social interactions are among the single most biologically important type of behavior in which humans and non-human primates engage. Life outcomes such as mental and physical health, status, material well-being, and reproductive success depend on an individual's ability to successfully navigate social interactions by deploying appropriate combinations of behavior. Direct observation of the fine structure of social interactions in free-ranging non-human primates can provide valuable insights into the neurobiological processes underlying social behavior by making it possible to quantify social phenotypes in terms of how they deploy particular combinations of behavior in social interactions. However, the complexity of social interactions makes the estimation of such phenotypes computationally and statistically challenging. I develop computational methods for modeling behavioral phenotypes in terms of the combinations of behaviors that characterize social interactions, and apply these methods to behavioral observations of rhesus macaques. I find that modeling the way animals combine social behaviors reveals novel information about animal social phenotypes that is lost when behaviors are considered individually, and I extend the model to directly incorporate biological and environmental covariates.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Neuroscience

First Advisor

Michael L. Platt

Subject Categories

Neuroscience and Neurobiology

MEASURING BEHAVIORAL PHENOTYPES FROM COMPLEX SOCIAL INTERACTIONS IN NON-HUMAN PRIMATES

Seth Madlon-Kay

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Michael L. Platt, Professor of Neuroscience

Graduate Group Chairperson

Joshua I. Gold, Professor of Neuroscience

Dissertation Committee

Joshua I. Gold, Professor of Neuroscience

Robert M. Seyfarth, Professor of Psychology

Danielle S. Bassett, Associate Professor of Bioengineering

Erol Akçay, Assistant Professor of Biology

MEASURING BEHAVIORAL PHENOTYPES FROM COMPLEX SOCIAL INTERACTIONS IN NON-HUMAN PRIMATES

ACKNOWLEDGEMENT

A full accounting of those to whom I am deeply indebted would fill several volumes, so I will thank just a few of those who contributed most directly to the work contained herein and my ability to see it through. My colleagues Weisong Ong and Geoff Adams provided intellectual stimulation and emotional commiseration in equal measures. Professor Lauren Brent made available to me the remarkable behavioral data set upon which all my work relies, and she, as well as Professor Robert Seyfarth, took it upon themselves to look after my education in primatology. The Carribean Primate Research Center and the efforts of the staff at Cayo Santiago made Lauren Brent's research, and by extension my own, possible. My dog Sherlock guarded me vigilantly throughout many late nights. My advisor Professor Michael Platt and committee chair Professor Joshua Gold showed endless patience and channeled my sometimes haphazard intellectual efforts in more fruitful directions. Finally, the support of my wife Ashley Walker makes it possible for me to do anything at all.

ABSTRACT

MEASURING BEHAVIORAL PHENOTYPES FROM COMPLEX SOCIAL INTERACTIONS IN NON-HUMAN PRIMATES

Seth Madlon-Kay

Michael L. Platt

Humans are innately social animals, and social interactions are among the single most biologically important type of behavior in which humans and non-human primates engage. Life outcomes such as mental and physical health, status, material well-being, and reproductive success depend on an individual's ability to successfully navigate social interactions by deploying appropriate combinations of behavior. Direct observation of the fine structure of social interactions in free-ranging non-human primates can provide valuable insights into the neurobiological processes underlying social behavior by making it possible to quantify social phenotypes in terms of how they deploy particular combinations of behavior in social interactions. However, the complexity of social interactions makes the estimation of such phenotypes computationally and statistically challenging. I develop computational methods for modeling behavioral phenotypes in terms of the combinations of behaviors that characterize social interactions, and apply these methods to behavioral observations of rhesus macaques. I find that modeling the way animals combine social behaviors reveals novel information about animal social phenotypes that is lost when behaviors are considered individually, and I extend the model to directly incorporate biological and environmental covariates.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1 : INTRODUCTION

Much of human life consists of a continuous series of social interactions, ranging from the everyday necessities of purchasing groceries from a store clerk, navigating a crowded sidewalk, or sharing meals and casual conversation with friends and family, to high-stakes encounters such as a job interview or a meeting with a potential romantic partner. Humans are typically able to fluently navigate such interactions by choosing an appropriate combination of actions for a given social context, and the ability to do so is critical to human well-being. Severe difficulty navigating social interactions can be considered pathological and is associated with neurological disorders such as autism spectrum disorder and schizophrenia (American Psychiatric Association, 2013; Barak & Feng, 2016; Burns & Patrick, 2007), while the inability to build and maintain social bonds leads to increased risk of a wide variety of poor mental and physical health outcomes (House, Landis, & Umberson, 1988; Steptoe, Shankar, Demakakos, & Wardle, 2013; Umberson & Karas Montez, 2010). In accordance with the importance of sociality to human life, many researchers from across the biological sciences, ranging from genetics to cognitive neuroscience, have over the last few decades increasingly directed their efforts towards understanding the biological basis of human social behavior (Insel, 2010).

However, for reasons logistical, ethical, and computational, understanding the biological processes underlying natural human social behavior is among the most difficult challenges of modern neuroscience. One major difficulty is collecting and analyzing ecologically relevant data. Research on human subjects is often, for logistical and ethical reasons, unable to directly observe and measure social interactions themselves, instead using indirect measures such as survey responses or demographic data that can collected inexpensively and non-invasively. Such measures may reflect the outcomes of social interactions, but they leave unobserved the underlying social interactions themselves. Laboratory research in humans is able to directly observe interactions, but necessarily takes place in artificial, low-stakes environments where only a relatively narrow slice of the human social repertoire can be

expressed and measured. While standard laboratory model organisms such as mice and flies permit direct and comprehensive observation of social interactions (see Robie et al. (2017) for a recent example), the more limited repertoires of social behavior in those species may limit the applicability of such research to human social behavior. Furthermore, such model organisms may differ from humans in the biological mechanisms associated with social behavior and live in laboratory environments with much less varied and dynamic social stimuli than the natural social environments of humans. Researchers rarely have opportunity to directly observe human social interaction in realistic environments. A consequence of these challenges is that we have few examples of biological explanations that causally and mechanistically link ecologically relevant human social behavior to a specific neurobiological or physiological process.

An alternative approach to the biological study of social behavior to use free-ranging populations of non-human primates (NHP), whose biology and social behavior are much closer to our own than animal models such as rodents or flies. Free-ranging NHP provide ecologically valid and biologically-relevant variability in social behavior that cannot be modeled in the laboratory or easily collected in humans. In particular, they provide an opportunity the observe the fine detail of social interactions and trace their relationship with the kinds of broader outcomes that can be measured in humans. For example, whereas in humans it is possible to assess, usually by self-report, the number and quality of a person's social relationships, NHP provide an opportunity to directly observe how relationships manifest as social interactions. This is important for the biological study of behavior because behavior at the level of the social interaction is closer to direct neurophysiological control than long-run summaries or outcomes of social interactions.

However, the complexity of social interactions presents serious methodological challenges. Social interactions between primates, both human and otherwise, are characterized not only by the use of behaviors at a given rate, but by the use of a particular combination of behaviors in sequence. The space of all possible ways in which individual behaviors can be

combined into patterns cannot be traversed using standard statistical methods. Accordingly, even research in NHP typically represents the behavioral phenotypes of animals in terms of summaries of long-run rates of behaviors, rather than in terms of the fine structure of social interactions. The second chapter of this thesis provides such an example of neurobiological research of social behavior using free-ranging NHP where social phenotypes are measured in terms of rates of individual behaviors.

The primary aim of this thesis is, first, to develop a methodological framework for quantifying the social phenotypes of non-human primates in terms of the combinations of behaviors that characterize social interactions. To accomplish this, I draw from methodological advances in machine learning for identifying complex but interpretable structure in high-dimensional data (Bishop, 2006; Murphy, 2012). Second, I investigate the usefulness of this framework in terms of what new information it can reveal about rhesus macaque social phenotypes compared to existing methods for quantifying social phenotypes that are standard in the animal personality literature. These efforts and their results are described in the third chapter. In the fourth chapter, I show how this methodological framework can be extended to include large quantities of data and behaviors, and directly infer relationships between the use of patterns of behavior at the individual level and biological and environmental covariates of interest.

In this work the model species used is the rhesus macaque (*Macaca* mulatta), due to their extensive use in lab and field research, neural circuitry which is homologous to that of humans, and complex social behaviors that contribute to biological success. The frequent use of rhesus macaques in laboratory, neurophysiological studies of behavior, including an increasing number in the domain of social behavior, makes it possible to draw links from the neural mechanisms of constrained laboratory behavior to the kinds of natural, unconstrained behavior that is our focus here (Pearson, Watson, & Platt, 2014).

A brief profile of rhesus macaques

The animal population whose behavior is examined throughout this thesis is the colony of rhesus macaques living on the island of Cayo Santiago, a 37-acre island off the southeastern coast of Puerto Rico. As described by one of the colony's founders, since its beginning the colony has been used as a resource for ecologically relevant natural social behavior, providing opportunities "not only to study animals as wholes but to study whole animals in their natural groupings in that environment which has been operating selectively on the species" (Carpenter, 1942).

Much of our basic knowledge of the nature of rhesus macaque social life derives from early research on Cayo Santiago by Carpenter and his colleagues (Rawlins & Kessler, 1986). Rhesus macaques are a despotic species with linear dominance hierarchies and social interactions characterized by strong dominance asymmetries. Lower ranking monkeys risk severe aggression if they do not submit to higher ranking animals (Thierry, 1985; Waal & Luttrell, 1989). Males and females live together in mixed social groups, and males emigrate from their natal group around the age of sexual maturity, while females typically remain in their natal group (Altmann, 1962). Among females, rank is inherited matrilinealy, with daughters tending to fall directly below their mothers in the dominance hierarchy. Multiple matrilines typically coexist within a single social group.

The colony on Cayo Santiago has been used to investigate the way physiological, hormonal, and genetic factors drive complex social behaviors, dating back to observations of Carpenter (1942) that the interplay between reproductive state and dominance status greatly impacted the nature of social interactions among male and female macaques (see Widdig et al. (2016) for a thorough review). More recently, Higham et al. (2011) has examined the role of oxytocin in female parental care, Brent et al. (2013) studied the relationship between serotonergic genotype and patterns of relationships among adult females, Trefilov, Berard, Krawczak, and Schmidtke (2000) found that serotonergic genotype impacted that age at which animals would leave their natal groups, and Charpentier, Prugnolle, Gimenez, and Widdig (2008)

investigated the relationship between genomic heterozygosity and rates of social behaviors.

Natural social phenotypes in humans and non-human primates

A major finding of the biological study of human behavior is that most human phenotypes that can be measured, including ones of social significance, are predictable on the basis of genetics, with most studies estimating phenotypic heritability in the range of 0.3 to 0.8 (Polderman et al., 2015; Turkheimer, 2000). Martin et al. (1986) found genetic influence on social attitudes ranging from jazz music to the death penalty. Numerous studies have found genetic influence on summaries of sexual behavior, such as the age of first date and lifetime number of sexual partners (see Harden (2014) for a general review). Similar genetic influence has been found in social behaviors such as political participation (Fowler, Baker, & Dawes, 2008), as well as general measures of social temperament such as the "Big Five" personality scales (Jang, Livesley, & Vernon, 1996). In recent years it has become possible to identify specific genetic variants associated with social behaviors and outcomes, such as educational attainment (Okbay et al., 2016; Rietveld et al., 2014).

Despite the ubiquity of genetic influence on behavior, we know little about the specific neurobiological mechanisms that give rise to complex social behavior. Turkheimer (2000) attributes the difficulty of identifying biological mechanisms as the problem of weak genetic explanation: though there exists some statistical relationship between a genetic or neurobiological variable and complex behavior, there is not necessarily any clear causal pathway that links the two. As Fowler et al. (2008) describes the problem, "there are simply too many genes and too many causal steps between genes and behavior to expect that genetic analysis will ever lead to improved understanding" of a complex behavior. For example, a personality rating on a survey represents a coarse summary of some large number of social interactions, which is then filtered through a process of retrospective evaluation as the survey taker decides how to answer the specific question on the form in front of them. In contrast, in certain animal models, it has for example proved possible to trace the expression of specific types of naturally-occurring social behaviors in rodents to the activation of particular neural circuits

linking sensory inputs to the hypothalamus, and even to induce different types of affiliative or agonistic interactions via neural activation (Kruk, 1991; Nelson & Trainor, 2007). By examining social behavior in non-human primates, we may hope to alleviate this problem by studying social behaviors that are closer to a direct neurobiological cause. If a molecular or neurobiological covariate is identified as predicting this response, the relationship between such a covariate and the underlying social behaviors of interest will not necessarily be clear; the covariate could act on the social behaviors, the evaluative process, the act of responding, or any combination of the above. If we are interested in social behavior, the first of these is more relevant than the second or third processes. In non-human primates it is possible to study complex, human-like social interactions directly.

However, research in NHP does not typically quantify phenotypes in terms of social interactions. As in human research, social phenotypes in NHP typically consist of long-run summaries or outcomes of social interactions. Most common methods of social phenotyping in NHP based on natural social behavior fall into one of three categories: observer ratings, rates of individual behaviors, and network measures which describe an animal's social relationships. The first category, phenotyping from observer ratings uses survey responses from humans, often caretakers, who rate the animals on how well they match a variety of descriptors such as "curious" and or "aggressive" (Freeman & Gosling, 2010; Stevenson-Hinde & Zunz, 1978). As with human personality research, this method relies on subjective descriptions of social behavior over long periods of time, and in some cases the surveys are explicitly modeled after personality scales used in human research (King & Figueredo, 1997). The second category, phenotyping based on average rates of behavior involves the observation of individual animals for fixed intervals of time, repeatedly over a period of months or years, during which the observer records the occurrences of various behaviors defined in a predefined ethogram (Altmann, 1974). The number of occurrences of each behavior are then converted into rates at the end of the study period (Anestis, 2005; Brent et al., 2014; Koski, 2011; Martin & Suarez, 2017; Seyfarth, Silk, & Cheney, 2012; Seyfarth, Silk, & Cheney, 2014). The third category, social network measures, are

derived directly from observed behavior, but unlike measures from the previous category they combine the occurrences of behaviors with the identities of an animal's interaction partners to describe the nature of the animal's social relationships. Often such measures focus on an animal's direct connections (that is, dyadic relationships) with others, such as the percentage of possible partners an animal interacts with or how evenly they distribute their grooming activity among their total set of grooming partners (Bitetti, 2000; Koski, 2011). However, some recent work has examined measures of indirect connectedness as well (Brent, 2015; Brent et al., 2013). Additionally, many studies estimate animal phenotypes from responses to some experimental task or stimulus (Araya-Ajoy, Mathot, & Dingemanse, 2015; Freeman & Gosling, 2010); however, as we are currently concerned with natural social behavior in unconstrained environments, we do not consider this category further.

Rates of social behavior and network measures are, unlike observer ratings, objective and derived directly from observed social interactions. However, both types of measures can be thought of as summary statistics describing many social interactions, which necessarily discard information about how particular social interactions unfold. Therefore, while NHP research provides a unique opportunity to study the fine structure of complex, human-like social interactions, existing methods for estimating social phenotypes do not take full advantage of this opportunity. The goal of the research presented here is to develop methods which capture more of the complexity of NHP social interactions, and to determine what novel information we can learn from such methods about variability in NHP phenotypes.

CHAPTER 2 : WEAK EFFECTS OF COMMON GENETIC VARIATION IN OXYTOCIN AND VASOPRESSIN RECEPTOR GENES ON RHESUS MACAQUE SOCIAL BEHAVIOR

Abstract

The neuropeptides oxytocin (OT) and arginine vasopressin (AVP) influence pair bonding, attachment, and sociality, as well as anxiety and stress responses in humans and other mammals. The effects of these peptides are mediated by genetic variability in their associated receptors, *OXTR* and the AVPR gene family. However, the role of these genes in regulating social behaviors in non-human primates is not well understood. To address this question, we examined whether genetic variation in the OT receptor gene *OXTR* and the AVP receptor genes *AVPR1A* and *AVPR1B* influence naturally-occurring social behavior in free-ranging rhesus macaques – gregarious primates that share many features of their biology and social behavior with humans. We assessed rates of social behavior across 3,250 hours of observational behavioral data from 201 free-ranging rhesus macaques on Cayo Santiago island in Puerto Rico, and used genetic sequence data to identify 25 *OXTR*, *AVPR1A*, and *AVPR1B* single-nucleotide variants (SNVs) in the population. We used an animal model to estimate the effects of 12 SNVs (n=3 *OXTR*; n=5 *AVPR1A*; n=4 *AVPR1B*) on rates of grooming, approaches, passive contact, contact aggression, and non-contact aggression, given and received. Though we found evidence for modest heritability of these behaviors, estimates of effect sizes of the selected SNVs were close to zero, indicating that common *OXTR* and AVPR variation contributed little to social behavior in these animals. Our results are consistent with recent findings in human genetics that the effects of individual common genetic variants on complex phenotypes are generally small.

Introduction

The neuropeptides oxytocin (OT) and arginine vasopressin (AVP) regulate social behaviors across a variety of mammalian species. In various non-human primate (NHP) species,

introducing exogenous OT into the central nervous system promotes affiliative social relationships and pair bonding behaviors (Smith, Ågmo, Birnie, & French, 2010; Snowdon et al., 2010), increases social interaction, and inhibits social aversion (Parr, Modi, Siebert, & Young, 2013). Vasopressin has been less studied in NHPs, but may play a role in promoting paternal care in tamarins (Kozorovitskiy, Hughes, Lee, & Gould, 2006). In humans, OT has been implicated in a wide variety of social behaviors, ranging from trust and altruism in economic games (Kosfeld, Heinrichs, Zak, Fischbacher, & Fehr, 2005) to eye contact and social attention (Auyeung et al., 2015), as well as a role in reducing anxiety (Bartz, Zaki, Bolger, & Ochsner, 2011). However, a common limitation of both human and NHP research into the role of OT and AVP is the reliance on laboratory tasks in relatively small samples, with correspondingly less variable and dynamic social stimuli than in natural environments. Humans and many NHP species live in large social groups where maintaining relationships and navigating hierarchies are crucial to biological success (Brent, Ruiz-Lambides, & Platt, 2017; Fedigan, 1983; Silk, Alberts, & Altmann, 2003; Steptoe et al., 2013). Yet to what extent laboratory findings regarding OT and AVP generalize to richer, more realistic social environments remains an open question. In this chapter, we attempt to address this question by comparing variability in social behavior in free-ranging rhesus macaques (Macaca mulatta) in a naturalistic setting, to variability in the genes that encode OT and AVP receptors, *OXTR* and AVPR.

The *OXTR* and the AVPR family of genes encode the OT and AVP receptors respectively. While *OXTR* is the only oxytocin receptor gene, three AVP genes encode three different vasopressin receptors: *AVPR1A*, *AVPR1B*, and *AVPR2*. *AVPR1A* and *1B*, but not *2*, are expressed in the brain (Freeman, Inoue, Smith, Goodman, & Young, 2014). Since OT and AVP share similar amino-acid sequences, they can bind to each other's receptors, albeit with different affinities (Freeman et al., 2014; Young & Flanagan-Cato, 2012). This structural commonality may contribute to similarities in the range and type of processes that the two neuropeptides mediate. Researchers have consistently cited *AVPR1A* as the AVPR gene most relevant to social functions (Freeman & Young, 2016). In humans, genetic variants of

AVPR1A are associated with behaviors ranging from altruism in an economic game (Israel et al., 2008) to self-reported partner bonding (Walum et al., 2008). *AVPR1B* variation, while not implicated as directly in social behavior, is associated in humans with stress responses (Keck et al., 2008) and mood disorders (Dempster et al., 2007). *OXTR* polymorphisms in humans have been linked to variation in a wide range of social behaviors and phenotypes, ranging from attachment style (Costa et al., 2009) and prosociality in economic games (Israel et al., 2009) to emotion perception and stress reactivity (Rodrigues, Saslow, Garcia, John, & Keltner, 2009) (see Ebstein, Knafo, Mankuta, Chew, and Lai (2012) for a more thorough review).

The role of variation in these genes is less studied in NHPs than in humans, though several studies have investigated an indel polymorphisms in the 5' flanking region of *AVPR1A* in the social behavior of captive great apes (Hopkins, Donaldson, & Young, 2012; Latzman, Hopkins, Keebaugh, & Young, 2014; Staes et al., 2015; Wilson et al., 2017). Hopkins et al. (2012), Latzman et al. (2014), and Wilson et al. (2017) each examined the relationship between the long versus short allele and personality, as measured by observer questionnaires, in captive chimpanzees (Pan troglodytes). Hopkins et al. (2012) found a sex-by-genotype interaction whereby "dominance" and "conscientiousness" personality traits differed between males and females, but only among animals carrying a copy of the long allele, while Wilson et al. (2017) found using the same personality dimensions as Hopkins et al. (2012), that the long allele predicted decreased extroversion but with no interaction with sex. Latzman et al. (2014), using a similar questionnaire-based data set, but a different decomposition into personality dimensions, also reported a sex-by-genotype interaction wherein males carrying the long allele were higher for "dominance" and "stability" personality dimensions. Staes et al. (2015) also looked at captive chimpanzee personality, but used observed rates of specific behaviors to assess personality rather than questionnaire ratings, and reported that the long allele was associated with total time spent giving and receiving grooming. Staes et al. (2016) reported a study of the same polymorphism in captive bonobos rather than chimpanzees, using both observer questionnaires and behavioral rate observations, and found the long

allele associated with higher "attentiveness" and lower "openness". To our knowledge only Staes et al. (2015) has examined *OXTR* in NHPs social behavior, which reported no effect of an intronic SNV.

In the present study, we drew on a large set of behavioral and genetic sequence data from the free-ranging rhesus macaque colony on Cayo Santiago Island off the coast of Puerto Rico. Rhesus macaques are an excellent model organism for studying biological and environmental influences on social behavior due to their extensive use in laboratory and field research, and complex social behaviors that are critical to their survival and reproductive output (Brent et al., 2017; Brent et al., 2013; Massen & Koski, 2014). The large macaque population and minimal human intervention on Cayo Santiago provide a unique opportunity to study the genetic influence of *OXTR* and AVPR variation in a naturalistic setting where social environment more closely resembles conditions in the wild, and social behavior can directly impact biological success. Demographic, life history, and pedigree data are also available for the Cayo population, which allows us to disentangle the effects of specific genetic regions from the influence of environment and overall genetic similarity. Though OT and AVP have been implicated in a variety of behaviors that include both social and non-social, here our interest lay specifically with social interactions between monkeys that indicate the quality and nature of their social relationships. Accordingly, we focused on rates of giving and receiving grooming, approaches towards another macaque (approach), being in non-grooming physical contact (passive contact), aggression that resulted in physical contact between macaques (contact aggression), and aggressive actions and threats that did not result in physical contact.

Methods

Study site

The studied population is a colony of rhesus macaques living on the island of Cayo Santiago, a 15-hectare island located 1km off the southeastern coast of Puerto Rico (18° N, 66°

W). This is a free-ranging, freely-breeding population with known pedigrees, rich data on life histories and fitness, and extensive genetic and observational data on behavior. The colony was founded in 1938 with a population of 409 Indian-origin rhesus macaques and is currently maintained by the Caribbean Primate Research Center (CPRC; University of Puerto Rico Medical Sciences Campus). The population as of July 2017 numbered 1571 animals self-organized into six different social groups. 537 of the animals are adults of age six or above, and 758 are juveniles between the ages of one and five. Researcher and caretaker intervention in the population is minimal. Animals in the colony are provided commercially available monkey chow daily and unlimited access to water. Animals are handled only during designated annual trapping periods, during which infants are tagged for identification. Despite the lack of immigration since its founding there is little evidence for high rates of inbreeding on Cayo Santiago (Blomquist, 2009; Widdig et al., 2016). All procedures described below were approved by the University of Puerto Rico's Institutional Animal Care and Use Committee (IACUC #A6850108) and adhered to the legal requirements of the United States of America and the American Society of Primatologists' Principles for the Ethical Treatment of Primates.

Genetic data

Animals were captured by CPRC staff and technicians during annual trapping procedures. Following capture, subjects were caged and anesthetized for blood draws using an intramuscular injection of ketamine HCl, 10mg/kg body weight. Blood was drawn by animal health technicians, and animals were released after full recovery from anesthesia. DNA was immediately isolated from blood in the field using commercially available QIAGEN extraction kits (QIAamp DNA Blood Mini Kit). Extracted DNA was stored frozen until shipment to the Genomics and Microbiology Research Lab at the North Carolina Museum of Natural Sciences, where libraries were prepared for next-generation sequencing, and a catalog of variants were genotyped. We used standard tools to identify single nucleotide variants (SNVs) by aligning sequence reads to the three genes of interest (*OXTR*, *AVPR1A*

and *AVPR1B*) from the two most recent published rhesus macaque reference assemblies (Gibbs et al., 2007; Zimin et al., 2014), as well as the current reference assembly (rheMac8 or Mmul_8.0.1). Our variant calling pipeline integrated read alignment using bwa-mem (Li et al., 2009), PCR duplicate removal using picard, as implemented within SAMtools (Li & Durbin, 2009) and simultaneous SNV discovery using GATK (McKenna et al., 2010). We excluded SNVs with minor allele frequency <0.05, where the minor (vs major) allele refers to the allele that is less (vs more) frequent in the sampled population. If genotype coverage depth fell below a minimum of two reads, then genotypes were imputed using standard default parameters in the software Beagle version 4.1 (Browning & Browning, 2016). All variants were annotated using the software SNPeff (Cingolani et al., 2012), and SNVs of interest were those predicted as having high, moderate, or low impact. Genotypes and their predicted impact results were then compared across all three reference assemblies. All reported genotypes were denoted using rheMac8 genomic coordinates. Next, designated missense variants were assessed for predicted functional impact (e.g., protein structure, protein stability, binding affinity, etc.) using the SNAP2 browser (Hecht, Bromberg, & Rost, 2015). Finally, human orthologues of the macaque SNVs were identified using KAVIAR and the UCSC Genome Browser (Glusman, Caballero, Mauldin, Hood, & Roach, 2011; Kent et al., 2002). We used dbSNP (build 150) to determine whether any human orthologues had known clinical significance (Sherry et al., 2001).

Nearby SNVs are often highly correlated to the level of redundancy, which can cause issues in interpreting and estimating phenotypic effects. Accordingly, we iteratively identified pairs of the SNVs with a correlation >0.9 and randomly removed one of the SNVs, repeating the process until no such pairs existed. Only the SNVs which survived this process were included in the behavioral analyses.

Behavioral data

The data is comprised of ten-minute focal samples (Altmann, 1974). The order in which animals were observed was semi-randomized to equalize the times of day and year of each

animal's observation periods. Observers recorded the times at which the monkey engaged in any behaviors specified by a rhesus macaque-specific ethogram consisting of both social and non-social behaviors (Brent, 2010). A total of 201 macaques (123 females, 78 males) were both represented in the behavioral data set and had genotype data available, and were used for the present study. The behavioral data used in this study were 19,501 ten-minute focal observations collected from adult (age $>$ 6 years) male and female macaques from five social groups, F, R, V, HH, and KK. Observational data was collected from group KK in 2014, from group F during 2011 through 2016 , from group V during 2015 and 2016, and from group HH and R during years 2014 and 2015, respectively. If an animal had an unusually small number of focal observations taken for their group in a given year, their focal observations from that year were removed from the data set. The threshold for removal was two standard deviations below the mean number of focal observations across animals for their social group in that year. The total number of focal observations per animal across all years ranged from 174 (approximately 29 hours) to 11 (approximately 1.83 hours), with an average of 79.99 focal observations per animal (approximately 13.16 hours).

The following behaviors were analyzed in this study:

1. Grooming: Running the hands or mouth through the hair of another monkey for at least 5 s.
2. Passive contact: Sitting or lying in physical contact with another animal without grooming.
3. Approach: One individual approaches another to within arms' reach (2 m) without physical contact, and remains within that distance for at least 5 s.
4. Contact aggression: Direct physical contact such as a bite, hit, push, or grab.
5. Noncontact aggression: A lunge, charge, or chase that does not result in direct physical contact, or a threatening gesture that entails some combination of staring, barks, head bobs, and opening one's mouth with covered teeth.

Each behavior except for passive contact was further divided into whether the focal animal performed the action or received the action from another animal, for a total of nine interaction types. Only social interactions involving another adult macaque were used in this study;

interactions with infant or juvenile macaques were not used.

Pedigree data

We obtained animal pedigrees from a long-term database maintained by the CPRC. From the founding of the population up through 1992, maternal identity was ascertained by behavioral observations, such as nurturing behaviors and lactation. For most macaques born after 1992, both maternal and paternal identity were ascertained genetically through the analysis of 29 microsatellite markers (Brent et al., 2013). In this study, maternity was known from genetics for 195 macaques (97%), while paternity was known from genetics for 197 macaques (98%). When maternity was not known from genetics, maternity assignments from behavior were used. The population pedigree was used to generate a kinship matrix across animals via R package kinship2. We multiplied each element of the kinship matrix by two to create the genetic covariance matrix in order to measure the heritability of social behaviors. Maternal identity was known on average for 6 (1.4 sd) previous generations, and paternal identity was known on average for 2 (0.8 sd) previous generations.

Data processing and model specification

We used individual 10-minute focal observations as the basic unit of analysis rather than aggregating those focal observations into rates of behaviors for each animal across longer periods of time. The motivation for this approach is that when animals have different numbers of focal observations being aggregated into a single rate, the rates of animals with fewer focal observations will be intrinsically more noisy and less precise than those of animals with more focal observations, and thus should be weighed less. By using individual focal observations as data points, the number of focal observations itself for a given animal in a given year provides the appropriate weighting.

We represented each focal observation in terms of the total amount of each behavior that occurred during that focal observation. For behaviors with durations, such as grooming, that amount corresponded to the total time spent engaged in that behavior, while for

events such as approaches, it was the number of times that behavior occurred during an observation. The distributions of behavior amounts across focal observations were highly right-skewed and zero-inflated for each of the behaviors examined in this study. Furthermore, some behavior amounts were continuous (e.g., amount of time grooming), while others were discrete (number of times aggressive acts occurred). These issues rendered ordinary linear regression inappropriate. Therefore, for each focal observation, each behavior was discretized into one of three ordered categories, or levels. The levels corresponded to a behavior not occurring at all (none), occurring at a low rate (low), and occurring at a high rate (high) during a given focal observation. Focal observations in which the behavior did not occur were assigned to the "none" category, and we divided the remaining observations into the "low" and "high" categories by performing a median split on the behavior amounts. Note that the median used for the median split was calculated using only the focal observations not in the "none" category. The end result is that each focal observation was represented as a vector of category labels (none, low, or high), one for each behavior. This approach preserved information about relative amounts of behaviors while avoiding problems arising from mismatches between an assumed distribution of observations (e.g. normal or poisson) and the true distribution.

We assessed the contribution of the genetic variants to social behaviors using multivariate ordered logistic regression. We included age (linear and quadratic), sex, dominance rank (linear and quadratic), age-by-sex (linear and quadratic) and rank-by-sex (linear and quadratic) interactions as fixed effects covariates in the model. Nonlinear terms were included for age because the effect of aging one year likely changes across the lifespan, and for dominance rank because the difference between low and middle-ranked macaques may not be the same as the difference between middle and high ranked macaques. Dominance rank was represented on an ordered categorical scale: low-ranking animals outranked less than 50% of their social group, medium-ranking animals outranked between 50% and 80%; and high-ranking animals outranked greater than 80%. All covariates were mean-centered, and the linear and quadratic age terms were orthogonalized against each other and z-scored.

Additive genetic effects, permanent environment effects, maternal effects, and the year and group during which the observation took place (that is, observations from each year-group pair being grouped together) were included as random effects. We defined both genetic and permanent environment effects as those associated with particular animals that were consistent across focal observations of the same animal, but they differed in whether the effects were correlated across animals. Additive genetic effects refer to animal-specific effects that were assumed to be correlated across animals according to their kinship, here calculated using the Cayo pedigree, while permanent environment effects were assumed to be independent across animals (Fisher, 1918; Kruuk, 2004). We also note that permanent environment effects are "permanent" in the sense that they are consistent within an animal across the full timespan that the animal was studied. Finally, maternal effects were effects consistent across all focal observations of animals with the same dam (Wilson et al., 2010). The magnitude of the additive genetic variance component relative to the other sources of variation determined the narrow-sense heritability of a phenotype.

The effects of genotypes on behaviors were modeled as random effects with a heavy-tailed distribution, details of which are described in the section below. The motivation for treating genotypic effects as random rather than fixed is to provide regularization and prevent false positives (Gelman, Hill, & Yajima, 2012). This approach is also consistent with the approach used by genomic prediction tools, which generally assume that for complex phenotypes, many genetic variants have some small effect that comes from a common distribution that is estimated directly (Yang, Lee, Goddard, & Visscher, 2011; Zhou, Carbonetto, & Stephens, 2013). Animal genotypes were coded as the number of minor alleles at each locus (Balding, 2006; Yang, Manolio, et al., 2011).

Regression model and fitting procedure

We used a multivariate ordinal logistic version of the animal model:

$$
\begin{aligned}
\eta_{i,j} &= x_i'\beta_j + z_i'u_j + s_{a(i)}'\lambda_j + g_{a(i),j}\,, \\
\mathrm{p}(y_{i,j} > 0|\eta_{i,j}) &= \mathrm{logit}^{-1}(\eta_{i,j} - \alpha_j^{(0)})\,, \\
\mathrm{p}(y_{i,j} > 1|\eta_{i,j}) &= \mathrm{logit}^{-1}(\eta_{i,j} - \alpha_j^{(1)})\,, \\
\mathrm{logit}^{-1}(x) &= \frac{1}{1+\exp(-x)}\,,
\end{aligned}
\tag{2.1}
$$

where $y_{i,j}$ is the level (0, 1, or 2) of behavior j that occurred during focal observation i, x_i is a vector of fixed effects covariates, and z_i is a vector of random effects covariates. The vector s_a is the vector of SNVs belonging to the animal a, while $a(i)$ is the focal animal followed during observation i. The parameters β_j, u_j and λ_j are the regression weights for behavior j associated with the fixed effects, random effects, and SNV effects respectively. The parameters g_a represent the overall additive genetic component to the phenotype of animal a, as in the traditional animal model (Henderson, Kempthorne, Searle, & von Krosigk, 1959; Kruuk, 2004). The parameters α_j, where $\alpha_j^{(0)} < \alpha_j^{(1)}$, are offsets that determine the baseline probabilities of each level of behavior j.

To avoid false positives when estimating the genetic effects, we regularized the effect estimates using a flexible sparsity-inducing prior:

$$
\begin{aligned}
\lambda_{j,l} &\sim \text{Student's t}(4, 0, \sigma_\lambda^2)\,, \\
\sigma_\lambda &\sim \mathrm{N}^+(0, 2)\,,
\end{aligned}
\tag{2.2}
$$

where $\lambda_{j,l}$ is the effect of a minor allele at locus l on behavior j. The genetic effects are given a Student's t distribution centered at zero with four degrees of freedom, with a width σ_λ that is estimated from the data and is given a normal prior truncated at zero. This and similar priors have been used frequently in the estimation of genetic effects and predicting genetic values for animal breeding (Meuwissen, Hayes, & Goddard, 2001; Resende et al., 2012; Zhou et al., 2013). This prior has the property that effects for which the evidence is weak will be penalized and pooled towards zero, preventing overfitting, but because of the

t-distribution's heavy tails, large effects for which there is strong evidence will be preserved. Finally, the fixed effects, random effects terms are given weakly-informative priors (Gelman, 2006):

$$\begin{aligned} \beta_j &\sim \mathrm{N}(0, 2)\,, \\ u_j &\sim \mathrm{N}(0, \sigma_j^2)\,, \\ g_j &\sim \mathrm{N}(0, \gamma_j^2 A)\,, \\ \sigma_j &\sim \mathrm{N}^+(0, 2)\,, \\ \gamma_j &\sim \mathrm{N}^+(0, 2)\,, \end{aligned} \tag{2.3}$$

where A is the relatedness matrix among animals, and σ_j^2 and γ_j^2 are the random effect variance components and the additive genetic variance components respectively. Note that although for brevity only one random effect covariance component is listed in the equations, separate variances were fit for the animal identity and observer identity random effects terms.

Following Davies, Scarpino, Pongwarin, Scott, and Matz (2015), Nakagawa and Schielzeth (2010), and Vazquez et al. (2009), we estimated the narrow-sense heritability of each behavior j using the equation $\gamma_j^2/(\gamma_j^2 + \sigma_j^2 + \pi^2/3)$ where σ_j^2 is the sum of the variance components associated with permanent environmental and maternal effects. Note that heritability as estimated here refers to heritability of the latent continuous variables underlying logistic regression, rather than of the discrete behavioral data itself. Further, because fixed effects are not taken into account, it is the heritability in a population of animals belonging to the same social group and sex, and of similar age and rank, and so on.

We estimated posteriors for the model parameters using Markov-Chain Monte-Carlo sampling sampling via the Stan software package (Stan Development Team, 2016). We ran three chains with 1000 samples each, discarding the first 200 iterations of each as burn-in, for a total of 2400 samples used for inference. Convergence was assessed using the Rhat metric

reported by the rstan package.

We report the estimated effects of a SNV on behavior as the percent change in the odds of a behavior occurring associated with having one more copy of the minor allele. This corresponds to $\exp(\lambda) - 1$ rather than the raw λ values themselves. The reported point estimates of all parameters, effects, and quantities of interest are posterior means. We estimated the phenotypic variance contributed to behavior j by all the SNVs together as $\lambda_j' SS' \lambda_j$, where S is the matrix of mean-centered genotypes.

Results

OXTR and AVPR variants

We identified a total of 25 SNVs of interest (6 *OXTR*, 13 *AVPR1A*, and 6 *AVPR1B*). Table 2.1 shows genomic coordinates and descriptive information for these variants. Of these 25 SNVs, eight were missense variants (see Table 2.2). Two of the missense variants (one in *AVPR1A* and one in *AVPR1B*) were predicted to impact the structure of the receptor, according to assessments in SNAP2. Eleven SNVs had known analogues in the human genome, though none of those SNVs had known clinical significance (see Table 2.3). After pruning for high linkage disequilibrium among the 25 SNVs (see Methods), we retained 12 SNVs for phenotypic analysis (3 *OXTR*, 5 *AVPR1A*, and 4 *AVPR1B*).

Heritability and repeatability of rates of social behaviors

Before examining effects of specific single-nucleotide variants on social behaviors, we first assessed both genetic and non-genetic variability in social behaviors across individuals. Figure 2.1 shows the proportion of variance accounted for by additive genetic variance (heritability); the variance accounted for by permanent environmental effects; and the variance accounted for by maternal effects.

The three variance component estimates were modest for all behaviors. We estimated that the largest and smallest additive genetic effects accounted for 3.8% (passive contact)

and 1.2% (noncontact aggression received) of total variance of their respective behaviors. Similarly the largest and smallest estimated permanent environment effects were 5.1% (contact aggression received) and 0.6% (approach given) of total variability, and the largest and smallest maternal effects were 3.1% (grooming given) and 0.5% (approach received). Posterior uncertainty for all three variance components were such that negligibly small variance contributions could not be ruled out for most behaviors. The explained variance had a greater than one-in-ten chance of being below 1% for all variance components and all phenotypes, with the exception that the additive genetic variance component explained $<1\%$ of the variability of approaches (given) with probability 0.06.

Some of this posterior uncertainty is due to the fact that additive genetic and permanent environment are effectively correlated because permanent environmental effects are independant per-animal effects, and animals are genetically identical to themselves. Similarly, the fact that many dams had only one offspring in the data set (201 macaques with 156 unique dams) resulted in a correlation between maternal and permanent environmental effects. This relationship made it difficult for the model to distinguish between one effect being large and the others small, or the reverse. However, the sum of the three variance components was better determined, as can be seen in Figure 2.1, from the fact that in several behaviors, the sum had smaller 95% credible intervals than any individual variance component. The sum of heritability, permanent environmental effects, and maternal effects is the "repeatability" of a trait within individuals; that is, variability that is consistent within animals across observations, but that is not explained by the demographic and environmental variables in the fixed effects or the observation period random effect. Repeatability contributed $>1\%$ of total variance with probabilities >0.95 for all nine behaviors.

Effect sizes and credible intervals for the fixed effects parameters are shown in Figure 2.2.

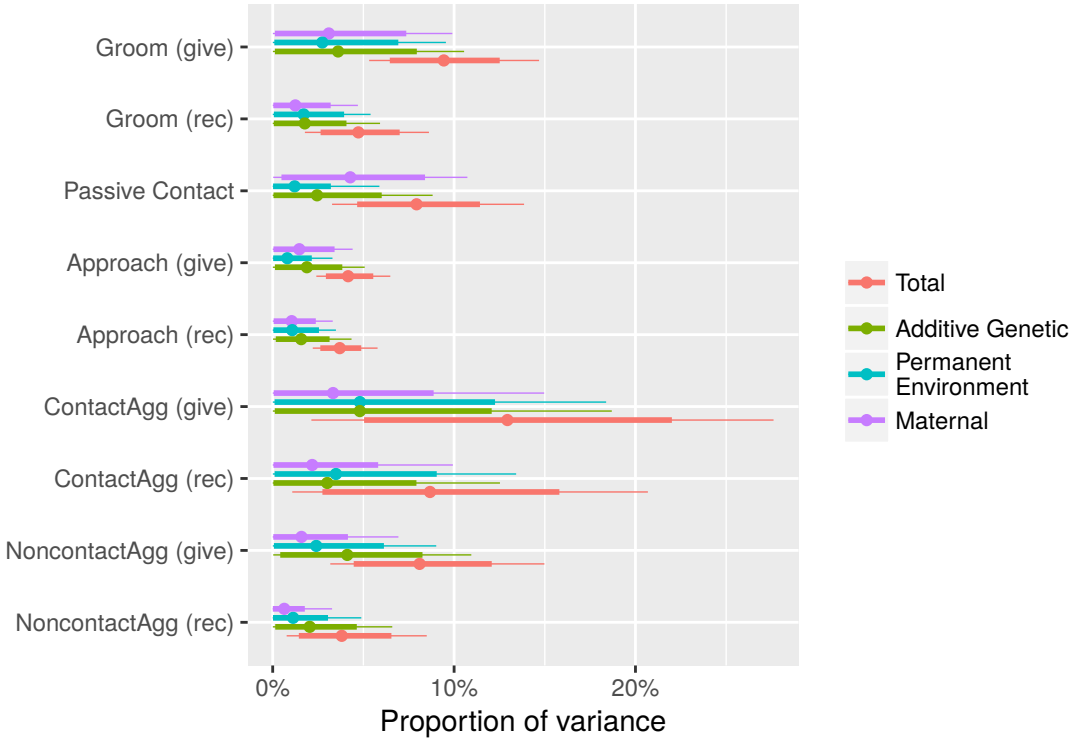

Figure 2.1: Proportion of residual variance attributed to additive genetic, permanent environmental, maternal effects, and the total proportion of residual variance explained by the three factors together. Points indicate posterior means, while thick lines and thin lines indicate 80% and 95% central credible intervals, respectively.

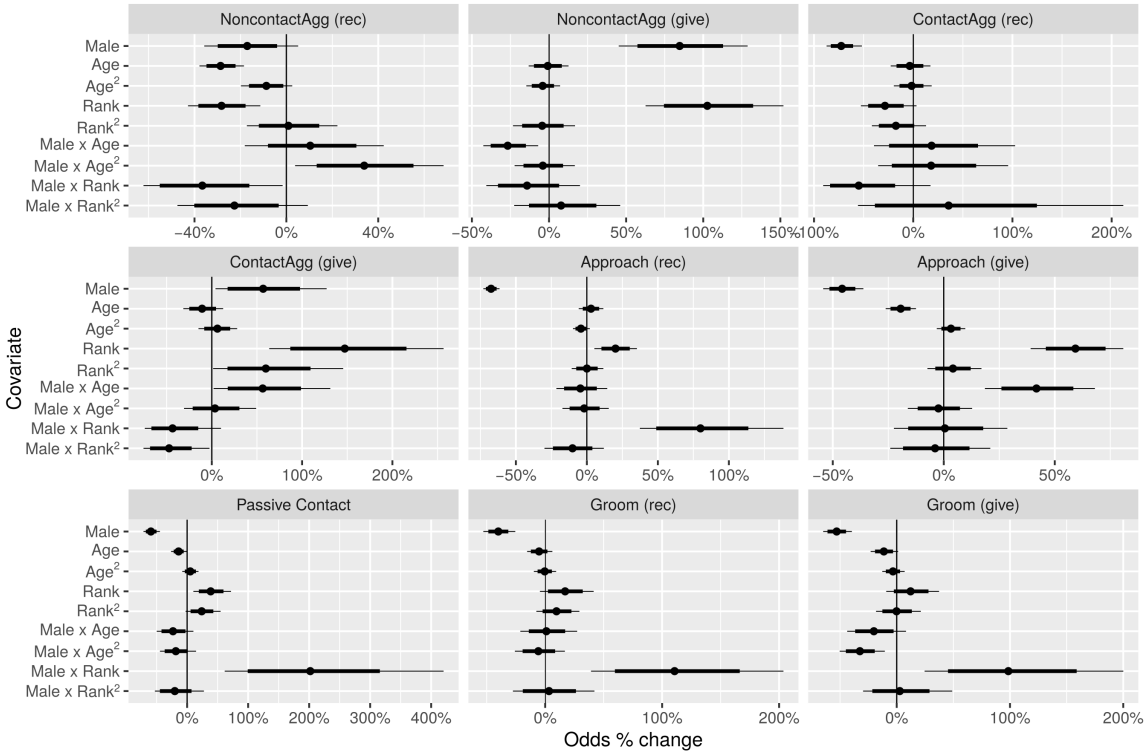

Figure 2.2: Effect sizes for fixed effects on social behaviors. Effect sizes are shown in terms of the percent change in the odds of a social behavior occurring during a focal observation associated with a unit increase in the dependant variable. Points indicate posterior means, while thick lines and thin lines indicate 80% and 95% central credible intervals, respectively.

Effects of SNVs on rates of social behaviors

The estimated effects on social behavior of oxytocin and vasopressin SNV minor alleles are shown in Figure 2.3. All effects were small, with the absolute average effect size estimated as a 0.2% change in the odds of a social behavior occurring, and the largest absolute effect estimated as a 1.7% change in odds. The 95% credible intervals (CI) included zero for all loci and all behaviors. Collectively, all SNVs together contributed between 0.03% and 0.04% of total phenotypic variation for each behavior. Beyond encompassing an effect size of zero, posterior distributions of the SNV effect sizes also indicated that the range of plausible effect sizes was small. Out of the 108 effects estimated (9 behaviors by 12 SNVs), 105 had 95% CIs that did not extend beyond an absolute effect size of a 6% change in odds. Of the remaining three effects whose CIs did exceed 6%, two were associated with *AVPR1A* missense variant chr11:62125302. The effects of this SNV were detected for approaches given (1.2% effect size, CI=[-1.7%, 8.7%]) and approaches received (1.6% effect size, CI=[-1.2%, 11.1%]).

Discussion

Our analysis of *OXTR* and AVPR genetic variants sought to measure the relationship between genetic variation in those genes and rates of spontaneous social interactions in a naturalistic setting in rhesus macaques, a highly social primate species. Our results are consistent with genetic variation in *OXTR* and AVPR having little to no influence on rates of social interaction. Though a number of previous studies found relationships between social behaviors and genetic variants in these genes, including in NHPs, this result is not entirely surprising and has several potential explanations.

First, many genetic associations identified in the human literature entailed analyses of behavior from laboratory tasks (Johansson et al., 2012; Knafo et al., 2008; Rodrigues et al., 2009) or from clinical phenotypes such as autism and mood disorders (Dempster et al., 2007; Israel et al., 2008; Lerer et al., 2008). It may be that genetic influences become attenuated in the more naturalistic social situations and non-clinical phenotypes studied here. Previous

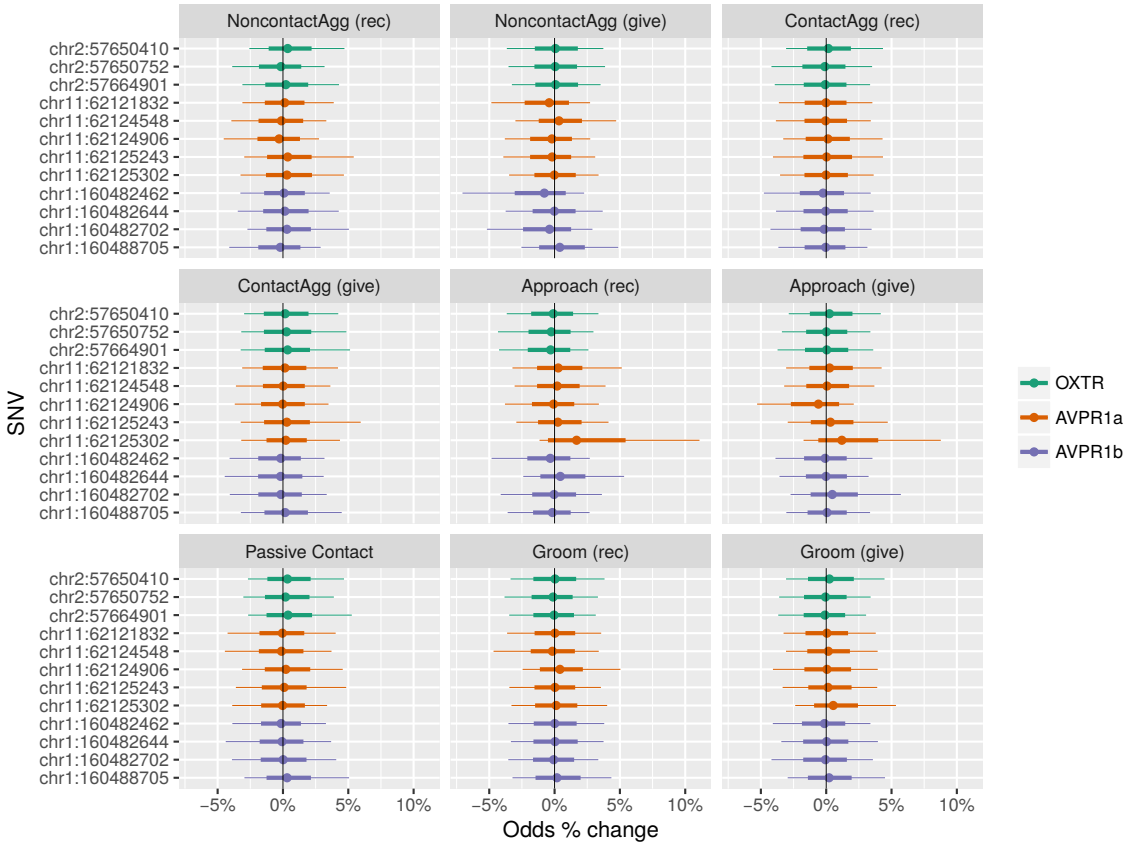

Figure 2.3: Effect sizes of *OXTR*, *AVPR1A*, and *AVPR1B* SNVs on social behaviors. Effect sizes are shown in terms of the additive effect of a minor allele on the percent change in the odds of a social behavior occurring during a focal observation. Points indicate posterior means, while thick lines and thin lines indicate 80% and 95% central credible intervals, respectively.

studies in great apes were more similar to ours in that they involved natural, non-clinical social behaviors, but to our knowledge these examined only captive populations in zoos and research colonies (Hopkins et al., 2012; Latzman et al., 2014; Staes et al., 2015; Wilson et al., 2017). Such settings are more constrained than Cayo Santiago in that animals cannot easily self-sort into distinct social groups, and human researchers and caretakers often intervene in reproductive success, health, access to resources for members of the population, and in order to prevent aggression that could cause serious injury or death. It is possible that in a naturalistic, unconstrained environment such as Cayo Santiago, environmental variability is larger and effectively "drowns out" genetic contributions.

Second, our results are consistent with recent findings that in general, complex behavioral and morphological phenotypes have a massively polygenic genetic architecture and individual variants have very small effects (Anney et al., 2012; Benjamin, Cesarini, Chabris, et al., 2012; Boyle, Li, & Pritchard, 2017; Chabris et al., 2013; Davies et al., 2011; Yang et al., 2015; Yang et al., 2010). Because effect sizes are typically small, historically both genome-wide and candidate gene studies of complex phenotypes with small sample sizes have low replication rates and are prone to false-positives (Chabris et al., 2012; Hart, de Wit, & Palmer, 2013; Ho et al., 2010; Ioannidis, Tarone, & McLaughlin, 2011; Siontis, Patsopoulos, & Ioannidis, 2010). We know of two reported replication failures of *OXTR* gene effects (Apicella et al., 2010; Munk, Hermann, El Shazly, Grant, & Hennig, 2016). Furthermore, two recent meta-analyses reported equivocal results regarding the influence of two heavily studied human *OXTR* SNVs on sociality (Bakermans-Kranenburg & van IJzendoorn, 2014; Li et al., 2015). Similarly, though several studies have implicated a 5' *AVPR1A* polymorphism in chimpanzee social behavior, the identified effects have been inconsistent. Hopkins et al. (2012) and Latzman et al. (2014) reported sex-by-genotype interactions and no main effects, whereas Staes et al. (2015) and Wilson et al. (2017) found only main effects of genotype on personality traits and no interactions with sex. While Staes et al. (2015) and Wilson et al. (2017) both reported effects of genotype on personality traits relating to prosocial and affiliative behaviors, the effects were in opposite directions,

with the same allele predicting higher prosociality in Staes et al. (2015) and lower in Wilson et al. (2017). While there may be unknown moderators that account for the differences between these studies, these inconsistent findings may also be the result of a lack of statistical power (Gelman & Carlin, 2014; Lemoine et al., 2016; Open Science Collaboration, 2015). Adding weight to this interpretation, we note that several of the papers listed above do not control for overall genetic relatedness within the population, do not adjust p-values for multiple comparisons or otherwise regularize effect estimates, or both, which can increase the likelihood of finding a false-positive genetic association. It may therefore be prudent to view *OXTR* and AVPR associations as preliminary, both in humans and in great apes, until they have been directly replicated.

Though we did not find evidence that *OXTR* and AVPR variants influenced social behavior, our results do suggest that social behaviors on Cayo Santiago have a modest additive genetic component. This is consistent with previous research on the same macaque population (Brent et al., 2013; Brent et al., 2014). This finding is also consistent with the theory that genetic influences on the social behaviors studied here are driven by small effects across large numbers of genetic polymorphisms (Fisher, 1918), however, it is worth noting that the magnitude of additive genetic effects relative to permanent environment and maternal effects was not well resolved in this study.

Our results do not indicate that all genotypic variability in the Cayo Santiago rhesus macaque population in *OXTR* and AVPR have small or no effects on rates of social interaction. Rare variants with very low MAF and de novo mutations are likely to have larger effects on complex phenotypes than common variants (Gratten, Wray, Keller, & Visscher, 2014; Neale et al., 2012), and because common variants are imperfectly correlated with rare variants and not at all with de novo variants (Eberle, Rieder, Kruglyak, & Nickerson, 2006; Speed, Hemani, Johnson, & Balding, 2012), those sources of genetic variability are not well captured by the SNVs genotyped in this study. Future research may profitably target rare rather than common genetic variants.

Gene	Position	Consequence Type	Read Depth	Mapping Quality	Reference Allele	Alternate Allele	Alternate Allele Frequency	Included in Behavioral Analysis
OXTR	chr2:57649859	missense variant	893	59.95	G	T	0.72	
OXTR	chr2:57649912	synonymous variant	917	59.97	A	C	0.34	
OXTR	chr2:57650182	synonymous variant	1117	60	C	T	0.34	
OXTR	chr2:57650410	synonymous variant	1018	60	C	G	0.48	Yes
OXTR	chr2:57650752	synonymous variant	1693	59.96	C	T	0.34	Yes
OXTR	chr2:57664901	synonymous variant	2717	59.99	A	G	0.28	Yes
AVPR1A	chr11:62121832	missense variant	1767	60	A	C	0.09	Yes
AVPR1A	chr11:62124427	synonymous variant	1830	53.03	C	T	0.91	
AVPR1A	chr11:62124548	missense variant	1912	60.01	C	A	0.36	Yes
AVPR1A	chr11:62124701	missense variant	1650	60	A	C	0.48	
AVPR1A	chr11:62124871	synonymous variant	1046	60	C	G	0.89	
AVPR1A	chr11:62124901	synonymous variant	1049	60	G	A	0.11	
AVPR1A	chr11:62124906	missense variant	1069	60	A	T	0.72	Yes
AVPR1A	chr11:62125186	synonymous variant	822	60	T	C	0.1	
AVPR1A	chr11:62125214	missense variant	947	60	A	G	0.09	
AVPR1A	chr11:62125231	synonymous variant	970	60	C	T	0.46	
AVPR1A	chr11:62125240	synonymous variant	925	60	T	C	0.1	
AVPR1A	chr11:62125243	synonymous variant	911	60	A	G	0.1	Yes
AVPR1A	chr11:62125302	missense variant	1027	60	G	A	0.48	Yes
AVPR1B	chr1:160482462	synonymous variant	1353	60	G	A	0.83	Yes
AVPR1B	chr1:160482464	missense variant	1290	59.97	C	T	0.89	
AVPR1B	chr1:160482644	synonymous variant	1288	59.97	G	A	0.87	Yes
AVPR1B	chr1:160482660	synonymous variant	1282	59.96	G	A	0.88	
AVPR1B	chr1:160482702	synonymous variant	1274	60	C	G	0.88	Yes
AVPR1B	chr1:160488705	synonymous variant	1615	59.99	T	C	0.44	Yes

Table 2.1: *OXTR*, *AVPR1A*, and *AVPR1B* SNVs

Though the relationship between social behavior, the molecules OT and AVP and their associated receptor genes, *OXTR* and AVPR has been studied extensively in laboratory settings in humans and captive animal populations, it is unknown to what extent those findings generalize to spontaneous behaviors in naturalistic environments. We examined this issue using an extensive behavioral and genomic data set from the free-ranging rhesus macaque population on Cayo Santiago, focusing on the relationship between *OXTR* and AVPR single nucleotide variants and social interactions related to the quality and kind of social relationships between animals. We found that the effects of SNVs in *OXTR* and AVPR on rates of social interactions were very small and possibly nonexistent, consistent with the idea that common genetic variants have generally weak effects on complex phenotypes.

Gene	Position	Amino Acid Change (refference/variant)	Amino Acid Position	Functional Effect	SNAP2 Score	Expected Accuracy
OXTR	chr2:57649859	Ala/Ser	6	no effect	-88	93.00%
AVPR1A	chr11:62121832	Ile/Val	419	no effect	-98	97.00%
AVPR1A	chr11:62124548	Ala/Glu	263	no effect	-70	82.00%
AVPR1A	chr11:62124701	Gln/Pro	212	no effect	-25	61.00%
AVPR1A	chr11:62124906	Met/Leu	144	effect	44	71.00%
AVPR1A	chr11:62125214	Asp/Gly	41	no effect	-38	66.00%
AVPR1A	chr11:62125302	Ala/Thr	12	no effect	-91	97.00%
AVPR1B	chr1:160482464	Ala/Val	84	effect	59	75.00%

Table 2.2: *OXTR*, *AVPR1A*, *AVPR1B* missense variants

Gene	Macaque Position	Human Genome Liftover Position	Human Genome Liftover rsID	Clinical Significance in Humans
OXTR	chr2:57649859	chr3:8768172		
OXTR	chr2:57649912	chr3:8768119	rs780323772	none known
OXTR	chr2:57650182	chr3:8767849	rs775129787	none known
OXTR	chr2:57650410	chr3:8767621	rs762128258	none known
OXTR	chr2:57650752	chr3:8767279	rs769535684	none known
OXTR	chr2:57664901	chr3:8752980	rs146441685	none known
AVPR1A	chr11:62121832	chr12:63147370		
AVPR1A	chr11:62124427	chr12:63149937		
AVPR1A	chr11:62124548	chr12:63150058	rs776846916	none known
AVPR1A	chr11:62124701	chr12:63150211	rs190242785	none known
AVPR1A	chr11:62124871	chr12:63150381	rs553995625	none known
AVPR1A	chr11:62124901	chr12:63150411		
AVPR1A	chr11:62124906	chr12:63150416		
AVPR1A	chr11:62125186	chr12:63150696		
AVPR1A	chr11:62125214	chr12:63150724		
AVPR1A	chr11:62125231	chr12:63150741		
AVPR1A	chr11:62125240	chr12:63150750		
AVPR1A	chr11:62125243	chr12:63150753		
AVPR1A	chr11:62125302	chr12:63150812		
AVPR1B	chr1:160482462	chr1:206116822		
AVPR1B	chr1:160482464	chr1:206116640	rs138075414	none known
AVPR1B	chr1:160482465	chr1:206116639		
AVPR1B	chr1:160482660	chr1:206116624		
AVPR1B	chr1:160482702	chr1:206116582	rs781803425	none known
AVPR1B	chr1:160488705	chr1:206110210	rs781813621	none known

Table 2.3: *OXTR*, *AVPR1A*, *AVPR1B* SNV human genome analogs

CHAPTER 3 : MODELING SOCIAL INTERACTIONS IN FREE-RANGING RHESUS MACAQUES

Abstract

Non-human primates are powerful model organisms for understanding complex social behaviors and the determinants of those behaviors among humans NHP behavioral phenotypes are usually estimated as a list of long-run average rates of individual behaviors. This limits the complexity that can be described in individual animals because it discards all information about the way animals deploy behaviors in coordinated patterns. We propose instead a representation of phenotypes in terms of the patterns in which behaviors co-occur, and develop two computational models by which this representation can be learned for individual animals even in small datasets. The second model borrows from Machine-Learning to describe behavior in terms of discrete strategies across all behaviors, which are shared by all animals in the population. We apply these methods to behavioral observations from a population of free-ranging rhesus macaques, and find that both models show much higher predictive power than the default of treating each behavior as independent. The behavioral strategy model further outperforms the correlative model, and identifies strategies which are interpretable and consistent with the existing literature on the complexity of NHP behavior. We also quantify the amount of information about individual animals we uncover through these methods, and find that representing individual phenotypes as patterns yields approximately 25% more information than representing behaviors as rates.

Introduction

A key characteristic of human social interactions is that we deploy particular combinations of behavior according to our social context and goals. A single behavior can be used as part of many different strategies for navigating social situations. For example, a ubiquitous human behavior such as speaking can take on distinct meanings and implications depending on whether it is combined with aggressive or relaxed body language. While one individual

might overall speak more than others, their long-run average amount of speaking may not by itself fully characterize the strategies they deploy in social situations. A more complete description of an individual's social phenotype should rely not just on the tallying of overall rates of individual behaviors, but on the particular combinations of behaviors they deploy in particular social interactions across different times and different contexts.

Measuring social phenotypes in terms of combinations of behaviors requires observing the fine details of natural social interactions. Research into the biological correlates of human social behavior rarely has direct access to behavior at such granularity. Alternative model organisms that do permit the direct study of complex social interactions are non-human primates (NHP). Like humans, many species of NHP live in large hierarchically-organized social groups and navigate their social environment by deploying particular combinations of behavior appropriate to their social context and goals. The field of animal personality has taken advantage of this opportunity to study the biological correlates of social interactions and their consequences (Gosling, 2001; Roche, Careau, & Binning, 2016; Freeman & Gosling, 2010). However, standard quantitative methods in animal personality research do not exploit the full complexity of NHP behavior because they represent behavioral phenotypes as long-run average rates of behaviors considered in isolation. In terms of the previous example, this simplification is analogous to considering how much an individual talks and how often they employ different types of body language, but not how that individual combines these behaviors into strategies for navigating their social environment. We hypothesize that the way animals combine social behaviors carries important information about the behavioral phenotypes of individual animals, and that a substantial portion of this information is lost when rates of behavior are considered in isolation. In the present work, we develop quantitative methods to test this hypothesis and do so in a large data set of observational data collected from free-ranging rhesus macaques living in a large social group. If the hypothesis is true, it indicates untapped potential for using rhesus macaques and other NHP species to identify the environmental and biological factors underpinning of the fine structure of social interactions.

As noted in Altmann (1962) regarding rhesus macaques in particular, NHP are able to choose particular combinations of behaviors "out of all available courses of action" in order to "maximize their gains and minimize their losses". For example, NHP have been found to choose grooming partners based on the benefits of affiliation compared to the cost of competition (Schino, 2001; Seyfarth, 1977; Seyfarth, 1980) and form coalitions in order to maintain or improve their social rank, using information about the relative ranks of their peers (Altmann, 1962; Maslow, 1936; Silk, 1999).

While this research has demonstrated the existence and importance of complex combinations of behavior at the level of NHP species, such patterns of behavior are not generally identified at the level of the individual animal. Examining the biological correlates of social behavior benefits greatly from an understanding of behavioral phenotypes at the level of the individual animal, because variability in social phenotypes across animals can then be compared with biology, e.g., genotypic variability. Standard methods for measuring the phenotype of an individual animal involves the observation of that animal over some periods of time during which the occurrences of various behaviors are recorded (Altmann, 1974). The estimate of the behavioral phenotype is these occurrences aggregated into rates of each behavior (Anestis, 2005; Capitanio, 1999; Koski, 2011; Martin & Suarez, 2017; Neumann, Agil, Widdig, & Engelhardt, 2013; Seyfarth et al., 2012; Seyfarth et al., 2014). Several such studies have found that behavioral phenotypes in NHP are heritable (Blomquist & Brent, 2014; Brent et al., 2014; Madlon-Kay et al., 2018; Staes et al., 2016), and some have also found effects of genetic variants in genes related to neuropeptide function in apes (Staes et al., 2015; Staes et al., 2016); though see Madlon-Kay et al. (2018). However, this analytic approach discards all information about the ways animals combine behaviors into strategies for dealing with particular social situations.

One reason prior research has considered behaviors independently rather than in combination is the computational and statistical difficulty of the latter. Animal personality research regularly deals with ten or more distinct behaviors, leading to 2^{10} possible combinations of

behaviors and requiring tens of thousands of observations of every animal in the studied population. To overcome this challenge, we utilize methods from Bayesian machine learning to develop computational methods for searching the space behavior combinations and identifying those combinations that are important for explaining observed behavior. These combinations are captured via a set of probabilistic strategies which are used across the population. During any given period of time an animal chooses one of these available strategies, with each animal's individual phenotype being described by the rates at which they use each of these possible strategies, as well as a global bias for each animal towards or against each behavior.

Here we test the hypothesis that behavioral phenotypes defined in terms of strategies involving combinations of beahaviors, as described above, provide novel and interpretable information about individual animals that is lost when rates of behavior are considered in isolation. We do so by applying the computational methods described above to extensive observational data collected from free-ranging rhesus macaques on the island of Cayo Santiago. First, we ask whether the strategies which we identify in the rhesus macaque population are consistent with our extant knowledge of NHP behavior. Second, we examine whether we find significant heterogeneity across animals in their use of those strategies, and whether that heterogeneity could be explained only by differences between animals in rates of individual behaviors. Finally, we develop an information-theoretic method for quantifying how much information we gain by considering behavioral phenotypes in terms of strategies versus rates of individual behaviors.

Methods

Study site

The studied population is a colony of rhesus macaques living on the island of Cayo Santiago, a 15-hectare island located 1 km off the southeastern coast of Puerto Rico. This is a free-ranging, freely-breeding population with known pedigrees, rich data on life histories and

fitness, and extensive genetic and observational data on behavior. The colony was founded in 1938 with a population of 409 Indian-origin rhesus macaques and is currently maintained by the Caribbean Primate Research Center (CPRC; University of Puerto Rico Medical Sciences Campus). The population as of July 2017 numbered 1,571 animals self-organized into six different social groups. 537 of the animals are adults of age six or above, and 758 are juveniles between the ages of one and five. Researcher and caretaker intervention in the population is minimal. Animals in the colony are provided commercially available monkey chow daily and unlimited access to water. Animals are handled only during designated annual trapping periods, during which infants are tagged for identification. All procedures described below were approved by the University of Puerto Rico's Institutional Animal Care and Use Committee (IACUC #A6850108) and adhere to the legal requirements of the United States of America and the American Society of Primatologists' Principles for the Ethical Treatment of Primates.

Behavioral data

The data is comprised of 10-min focal samples (FS) (Altmann, 1974). The order in which animals were observed was semi-randomized to equalize the times of day and year of each animal's observation periods. Observers recorded the times at which the monkey engaged in any behaviors specified by a rhesus macaque-specific ethogram consisting of both social and non-social behaviors (Brent, 2010). For this project we analyzed behavioral data collected from adult females of the largest and best-studied social group on the island, group F, during the years 2013 through 2016. This group constituted 94 animals, with an average of 19.7 (9.4 hour sd) hours of focal samples per animal. Because we are interested specifically in idiosyncratic variability in behavioral phenotypes that is not directly explainable by known situational factors, we excluded the highest ranking animals from the analysis, for a total of 80 animals. If an animal had an unusually small number of focal samples taken for their group in a given year, their focal samples from that year were removed from the data set. The threshold for removal was two standard deviations below the mean number of focal

samples across animals for their social group in that year.

Our analysis included eleven behaviors. Of those, three were self-directed:

- Self-directed behaviors
- Autogrooming
- Scratching
- Traveling
- Feeding

Five behaviors were affiliative:

- Groom (higher-ranked partner)
- Approach (higher-ranked partner)
- Groom (lower-ranked partner)
- Approach (lower-ranked partner)
- Passive physical contact

Three behaviors were aggressive or agonistic:

- Aggression
  - Threats, e.g. barks, head bobs
  - Non-contact, e.g. lunges
  - Contact, e.g. bites, grabs
- Submission
- Avoidance

Only interactions with adult female conspecifics were counted.

We controlled for various covariates by including fixed-effects that influence the rates of each behavior independently from the animal's identity. The fixed effects were the year of the focal sample coded as dummy variables and ordinal dominance rank. Rank was

measured on an ordinal scale: low-ranking animals outranked less than 50% of their social group, medium-ranking animals outranked between 50% and 80%, and high-ranking animals outranked greater than 80%.

Covariates were standardized such that the intercepts corresponded to population averages rather than baseline groups using the R package "standardize" (Eager, 2017).

Data representation

The basic unit of our behavioral analysis is the focal sample (FS). In order to render the analysis of behavioral combinations tractable, we make two major simplifying assumptions. First, we treat all time points within a single FS as equivalent; all events of the same behavior are treated as equivalent regardless of the relative or absolute timing of the event within an FS. Second, we also treat behavior occurrence within a FS as binary, representing them as either occurring or not, even if they occur more than once during a FS.

Formally, we represent FS as the random variable $Y \in \{0,1\}^D$, and denote a particular FS as the data vector $\mathbf{y}$. This is a D-dimensional binary vector where D is the number of behaviors considered. With D behaviors, Y can take 2^D possible values, corresponding to 2^D different combinations of behaviors. In the following section, we use $P(Y)$ to denote a probability distribution over all 2^D combinations and $P(Y|A)$ to denote the probability distribution associated with a specific animal A.

Models for the use of combinations of behavior in social interactions

With unlimited data, it would be possible to estimate directly for each animal the probability of exhibiting each of the 2^D combinations of behavior. With the modest amounts of data available from NHP populations, this is not feasible, so we instead rely on a model by which we may describe the probabilities of each combination using many fewer free parameters. At the heart of the model is a repertoire of strategies which exist in the population. A strategy consists of a probability associated with each behavior, which determines how likely

an animal is to exhibit that behavior while using that strategy. Because strategies are probabilistic in nature, many individual combinations of behaviors are compatible with a single strategy. For example, consider a strategy under which grooming and feeding have high probabilities, threats have low probability, and all other behaviors have moderate probabilities. Such a strategy is consistent with all combinations of behavior in which grooming and feeding occur and threats do not, regardless of what other behaviors occur.

Every animal has a unique probability distribution over strategies, which describes how often each animal uses each strategy. An intuitive interpretation of this aspect of the model is that at the beginning of a FS, the focal animal flips a K-sided die. Whichever side the die lands on determines which of the K strategies the animal uses for the duration of the FS, which in turn determines the combination of behaviors the animal will deploy during that FS. While the K strategies are shared across all animals, each animal has a unique die with different probabilities of landing on any given side compared to any other animal.

The model takes the form of a probit mixed-effects regression:

$$\begin{aligned} p_{Y|A}(\mathbf{y}_i|a_i) &= \prod_{d=1}^{D} \Phi(\eta_{i,d})^{y_{i,d}} (1 - \Phi(\eta_{i,d}))^{1-y_{i,d}} \,, \\ \eta_{i,d} &= \sum_{k=1}^{K} \alpha_{k,d} \mathbf{1}_k(z_i) + u_d^{(a_i)} + X_i' \beta_d \,, \\ \mathbf{u}^{(a)} &\sim \mathcal{N}(0, \Sigma) \,, \end{aligned} \tag{3.1}$$

where Φ is the cumulative distribution function of the standard normal distribution, $z_i \in \{1, \ldots, K\}$ indicates which of the K behavioral strategies animal a_i used on focal sample i. The parameters $\alpha_{k,d}$ are the propensities to engage in behavior d while using strategy k. The random effects $u_d^{(a)}$ represents an animal's global bias towards or against each behavior d, which is active across all of that animal's focal samples regardless of the behavioral strategy used.

Each animal's rate of using each strategy is determined by a mixed-effects multinomial

logistic regression:

$$
\begin{aligned}
p(z_i = k) &= \frac{\exp \theta_k^{(a_i)}}{\sum_{k'=1}^{K} \exp \theta_{k'}^{(a_i)}}\,, \\
\theta_k^{(a_i)} &= \mu_k + \epsilon_k^{(a_i)}\,, \\
\epsilon_k &\sim \mathcal{N}(0, \sigma_k^2)\,,
\end{aligned}
\tag{3.2}
$$

where μ_k determines the overall frequency of strategy k across the population, and the per-animal random effects ϵ determine how animal a's rate of using each strategy deviates from the population average.

The number of strategies available in the population is fixed at K. To determine an appropriate number of strategies we fit multiple versions of the BS model with varying K, denoting a model with exactly K strategies as BSK. For examination of the contents of the BS model strategies, we select the least complex model with performance within one standard error of the best-performing model (Hastie, Tibshirani, & Friedman, 2009). Here, less complex corresponds to smaller K.

To address the concern that the particular structure of the BS model might be a poor match for the macaque behavioral data, and that therefore conclusions drawn on the basis of it may be misleading, we also compare the BS model to an existing model for describing combinations of binary variables. Specifically, we adapt the quadratic exponential binary model (Cox, 1972; Cox & Wermuth, 1994), in which the probability of a combination of behaviors is determined by main effects of each behavior and second-order interactions between them. The main effects and interactions here are analogous to the means and correlations among a set of continuous variables. Intuitively, the BS model identifies a small number of high-order interactions between behaviors, whereas the correlative model uses all possible second-order interactions.

Formally, the model is given by

$$
\begin{aligned}
p_{Y|A}(\mathbf{y}|a_i, X_i) &= \Delta^{-1} \exp(\eta_i)\,, \\
\eta_i &= \sum_{d=1}^{D} y_d^* (\alpha_d^{(a_i)} + X_i' \beta_d) + \sum_{d_1=1}^{D-1} \sum_{d_2>d_1}^{D} y_{d_1}^* y_{d_2}^* \gamma_{d_1 d_2}^{(a_i)}
\end{aligned}
\tag{3.3}
$$

where y_d is the dth element of the FS $\mathbf{y}$, a_i is the focal animal observed during the ith FS, α and γ are the main effects of and interactions between behaviors d, and Δ is the normalizing constant, and X_i sample-specific covariate vectors. We follow Cox (1972) in using the centered transformation of the data, $y_d^* = 2y_d - 1$.

This model represents the behavioral phenotype of an animal a using $D + D(D-1)/2$ parameters. Rather than estimating these parameters independently for each animal, we use a hierarchical formulation of the model:

$$
\begin{aligned}
\boldsymbol{\alpha}^{(a)} &\sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)\,, \\
\gamma_{d,d'}^{(a)} &\sim \mathcal{N}(\nu_{d,d'}, \sigma_{d,d'}^2)\,, \\
\sigma_{d,d'} &\sim \mathcal{N}(0, \tau^2)\,,
\end{aligned}
\tag{3.4}
$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_d]'$ is the vector of main effects. This formulation pools each animal's main effects and interactions towards the population means μ and ν respectively, according to the inferred population (co)variances Σ and σ^2 (Gelman & Hill, 2007).

Prior specification and inference

For the correlative model, we use weakly informative priors on the hyperparameters (Gelman, 2006; Gelman, Jakulin, Pittau, & Su, 2008):

$$
\begin{aligned}
\mu &\sim \mathcal{N}(0, 25)\,, \\
\beta &\sim \mathcal{N}(0, 25)\,, \\
\Sigma &= (\mathrm{I}\boldsymbol{\psi})\Omega(\mathrm{I}\boldsymbol{\psi})\,, \\
\Omega &\sim \mathrm{LKJCorr}(2)\,, \\
\psi &\sim \mathcal{N}(0, 1)\,, \\
\nu &\sim \mathcal{N}(0, 1)\,, \\
\tau &\sim \mathcal{N}(0, 1)\,,
\end{aligned}
\tag{3.5}
$$

where the correlation matrix Ω is distributed according to the LKJ distribution described in Lewandowski, Kurowicka, and Joe (2009). Inference was performed via the No-U-Turn Sampler algorithm (NUTS) for Hamiltonian Markov Chain Monte Carlo (MCMC) in the RStan software package (Carpenter et al., 2017). We used four chains of 400 samples each and a burn-in period of 150 samples and assessed convergence using the Rhat metric and visual inspection of the hyperparameter chains.

For the behavioral strategy model we performed inference via a custom Gibbs sampler implemented in the Julia programming language (Bezanson, Edelman, Karpinski, & Shah, 2017) with the following priors:

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0, 10)\,, \\
\beta &\sim \mathcal{N}(0, 5)\,, \\
p(\Sigma) &\propto |\Sigma|^{-(1+D)} \prod_{d=1}^{D} (2\Sigma_{d,d}^{-1} + 1)^{-(1+D/2)}\,, \\
\mu_k &\sim \mathcal{N}(0, 5)\,, \\
\sigma_k^2 &\sim \mathrm{InvGamma}(0.5, 0.05)\,.
\end{aligned}
\tag{3.6}
$$

Gibbs sampling was used because we found NUTS was unable to generate samples from the strategy model in a reasonable timeframe. We render the probit regression layer amenable

to Gibbs sampling by using the latent variable formulation from Albert and Chib (1993). To permit Gibbs sampling from the logistic distribution over z that determines individual strategy utilization, we use the Polya-gamma latent variable representation derived in Polson, Scott, and Windle (2013). For the covariance matrix Σ we used the prior described in Huang and Wand (2013), which is conjugate scale-mixture of Inverse-Gamma distributions and permits independent estimation of the variances and correlations as well inference by Gibbs sampling. Our implementation mirrors the algorithm developed for logistic topic modeling in Chen, Zhu, Wang, Zheng, and Zhang (2013).

We initialized the Gibbs sampler based on maximum likelihood estimates of a simplified version of the BS model. The simplified model included only the α_k terms and treated strategy utilization as fixed across the population. Estimation was performed from 100 random starting points, and the fit parameters with the highest likelihood were used to generate the initial values of z for each FS. Specifically, each z was sampled from a multinomial distribution over the K strategies, with the probabilities being the normalized likelihoods of each observation. The Gibbs sampler was run for 40k iterations, with the first 4k iterations discarded as burn-in, with convergence assessed by visual inspection of the chain and split-half Rhat on the hyperparameters. The Gibbs sampler for the strategy model was found to be less efficient than HMC inference via Stan for the correlative model, necessitating many more samples be taken.

We assessed the predictive power of each model using the Widely Applicable Information Criterion (WAIC), which approximates leave-one-out cross validation similarly to Akaike Information Criterion (AIC) (Vehtari, Gelman, & Gabry, 2017; Watanabe, 2010) but unlike AIC is applicable to Bayesian posteriors rather than maximum-likelihood point estimates. The use of WAIC enables comparison of models through their out-of-sample predictive power without incurring the severe computational cost of When used for model comparison, WAIC magnitudes have the same interpretation other information criteria such as AIC and BIC. Conventionally, an absolute difference of 6 or more between models providing strong

evidence in favor of the better model roughly analogous to statistical significance at the $\alpha = 0.05$ level, and a difference of 10 or more being very strong evidence corresponding to the data being over 100 times more likely under the better model than the alternative model (Raftery, 1995).

Quantifying information gained using population mutual information

Mutual information (MI) is a general measure of the association between two random variables (Shannon, 1948). Intuitively, it corresponds to the reduction in uncertainty in one variable when another variable is taken into account, or in other words the explanatory power of one variable on another. Importantly, we can use MI to quantify the diversity of behavioral phenotypes in a population, which we will refer to as the population MI (PMI).

Specifically we can calculate PMI as the difference between the uncertainty of behavior at the population level and the average uncertainty for individual animals:

$$\mathrm{H}(Y) - \mathrm{H}(Y|A)\,, \tag{3.7}$$

where $\mathrm{H}(Y) = -\sum_{\mathbf{y}\in Y} p(\mathbf{y}) \log p(\mathbf{y})$ is the Shannon entropy of the population distribution, and $\mathrm{H}(Y|A) = \mathrm{E}_{a\in A}[-\sum_{\mathbf{y}\in Y} p(\mathbf{y}|a) \log p(\mathbf{y}|a)]$ is the conditional entropy of behavior given that Y is being observed from animal a.

The most explanatory power that a source of biological or environmental variability could possibly have on behavioral phenotypes would be to explain entirely the differences between individual animals. Therefore PMI constitutes an upper bound on the MI between behavior and any source of variability that may explain differences in behavioral phenotypes between animals. PMI shares this characteristic with repeatability, the more traditional measure of behavioral diversity used in animal personality and ethology generally (Dingemanse & Dochtermann, 2013; Nakagawa & Schielzeth, 2010). However, repeatability is defined in terms of the variance of a single behavior, and has no immediate generalization to multiple

behaviors that can co-occur in arbitrary patterns.

We can make the connection between PMI and repeatability explicit by noting that both are examples of a general class of measures of explained dispersion:

$$\frac{\mathrm{D}(Y) - \mathrm{E}_X[\mathrm{D}(Y|X)]}{\mathrm{D}(Y)}\,, \tag{3.8}$$

where D is some function that measures dispersion (Efron, 1978; Haberman, 1982; Theil, 1970). For repeatability that function is variance, while for PMI that function is entropy.

Results

Behavioral strategies in the Cayo population

At the population level, we find that capturing combinations of behavior via the behavioral strategy model significantly improves predictive power compared to a conventional analysis that does not account for combinations of behavior (Figure 3.1a). The simplest model, BS1, only permits a single strategy and so does not take into account combinations of behaviors. Because in this model animals are assumed to differ only in their constant biases for or against behaviors, this model is equivalent to conventional analyses that describe animal phenotypes in terms of rates of individual behaviors. All BS models with multiple strategies outperform BS1. The model with the best performance was BS8 (WAIC 57838, SE 350), while BS6 was the simplest model whose performance fell within one standard error of the minimum (WAIC 57926, SE 351). We therefore use BS6 in subsequent analyses of the behavioral strategy model.

Examining behavioral strategies estimated under BS6 (Figure 3.2), we find that they describe strategies of varying riskiness in foraging and socializing. Animals, on average, used strategies S1-3 on 87% (SE 1%) of FS, with use rates being roughly even among those top three strategies. Strategy 2 is a low-risk, low-reward strategy in which the animal experiences

low rates of feeding and both affiliative and agonistic interactions. Strategies 1, 3, and 5 index foraging with varying riskiness; S1 and S5 have high feeding rates at the cost of high rates of agonistic interactions, and S2 has lower rates of agonism at the cost of lower feeding rates. Strategy 5 is distinguished from the more common S1 by even higher rates of aggression and a lack of travel, possibly indicating fierce competition or defense of resources versus the opportunistic foraging of free resources. This distinction between competition over localized resources versus foraging for dispersed resources has been previously observed in Japanese macaques (Mori, 1977) among other NHP (Hill, 1999). S4 and S6 are different social strategies – the former socializing with animals of similar or lower rank, and the latter socializing with higher-ranking animals. Consistent with the competitive grooming theory described in Seyfarth (1977) and Schino (2001), the latter occurs less often than the former (posterior means 6% and 1%, SE 0.7% and 0.3%), and involves high rates of all agonistic behaviors.

Individual variability in strategy use

We have shown that accounting for the use of combinations of behavior improves our predictive power for rhesus macaque behavior at the population level. However, the question remains as to whether individual animals differ in their use of combinations of behavior. First, we compare the predictive power of BSK to a submodel in which all animals are assumed to have the same rates of using each strategy. We implement this submodel by fixing the parameter σ_k^2 to zero for each strategy k. Under this submodel, the only differences in the behavioral phenotypes of each animals can be described by differences in the global biases towards individual behaviors (the parameters u), regardless of the strategy being used. For all models other than BS2, we find strong evidence in favor of the full model versus the submodel (Figure 3.1b). The smallest difference in performance among BS4-8 was 214.6 (56.7 SE). Second, we examine the inferred variability in strategy use across the Cayo population. We find substantial heterogeneity in the rates at which different animals use each strategy. The three most frequently used strategies, S1-3, have posterior mean

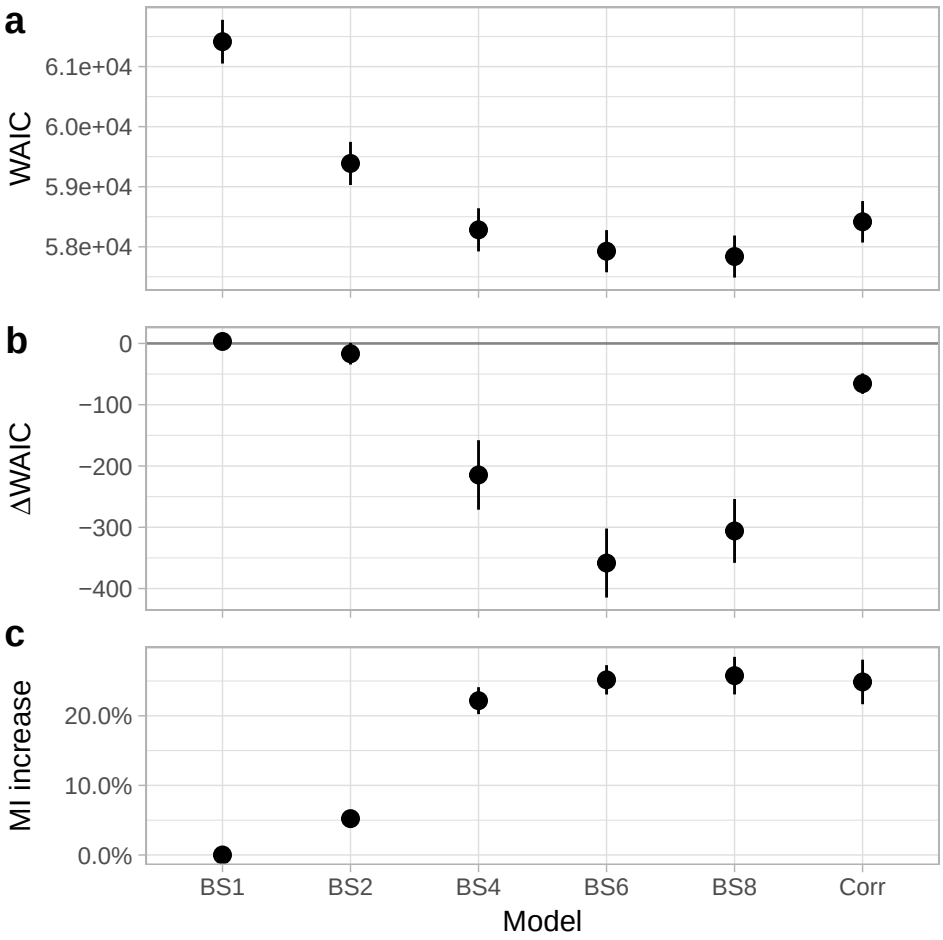

Figure 3.1: Quantitative comparisons of models for social interactions. a) Widely-applicable information criterion (WAIC) for models BSR1-8 and the correlative model. Lower is better. b) Difference in WAIC between each model and it's equivalent submodel where phenotypes are restricted to differ only in their rates of behavior. c) Percent increase in PMI under from using the full phenotype of each animal, versus reducing the phenotypes to rates of individual behaviors considered separately as in conventional analyses. Dots and lines are population means and standard errors in a) and b), posterior means and standard deviations in c).

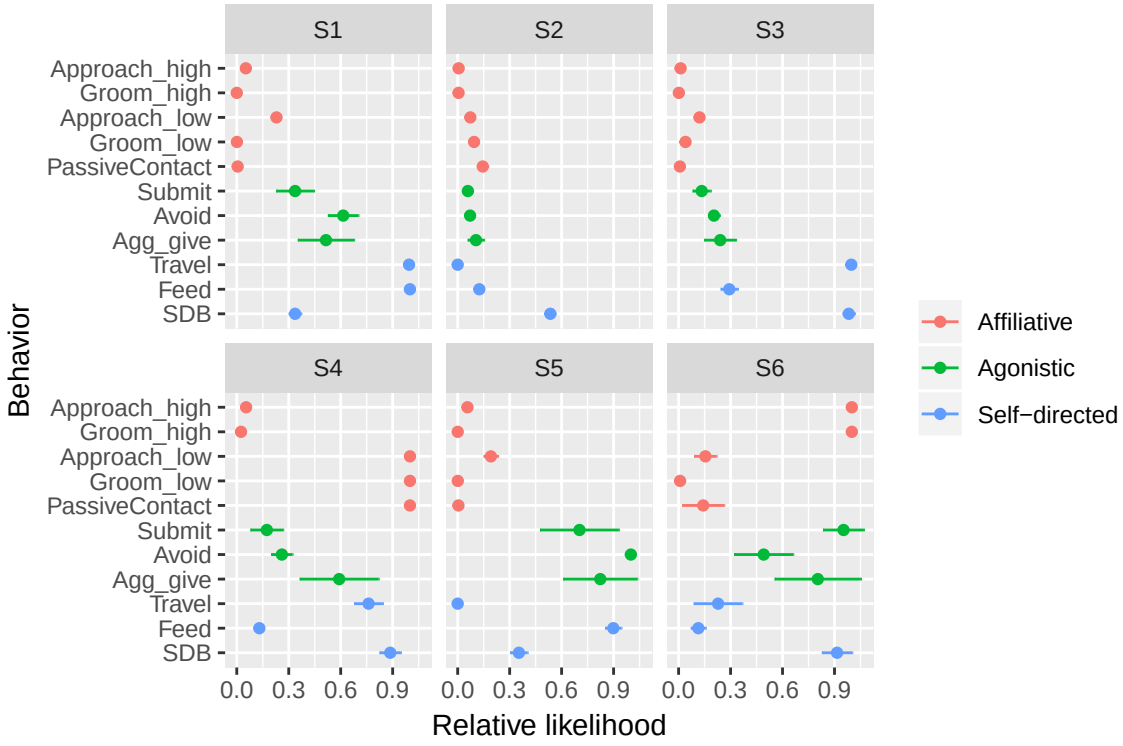

Figure 3.2: Relative probabilities of socially-relevant behaviors occurring in each of the strategies identified by BS6. Behavior probabilities are normalized such that 1 is the maximum probability of a given behavior across all strategies. Dots and lines represent posterior means and standard errors.

coefficients of variation near 18%, while strategies S4-6 have higher coefficients of variation at 60% or above, albeit estimated less precisely due to the less frequent occurrence of these strategies.

Information gained by the estimation of strategies

While the previous analyses have shown that animals differ measurably in their use of combinations of behavior, the relative importance of these differences to understanding overall differences in behavior between animals is still unclear. By modeling the way animals deploy combinations of behavior, how much more have we learned about their individual behavioral phenotypes than we would have from looking at rates of individual behaviors? To answer this question we estimate the mutual information in the population (PMI) between observed animal behavior and animal identity. PMI effectively quantifies the amount of behavioral diversity between individual animals compared to the total diversity of behavior observable within population. PMI also represents the upper-bound for how much information any environmental or biological factor can carry about behavior, because no determinant of behavior could do better than perfectly predicting the behavioral phenotypes of animals in the population.

Our present interest is not in the estimated PMI *per se*, but rather the increase in PMI from modeling the use of combinations of behaviors. We calculated, under each model, the PMI that we would expect from an analysis of rates of behavior. Specifically, for each posterior sample we discard all information about the behavioral phenotypes of all animals other than the rates of behaviors, calculate PMI as though each behavior occurred independently at those rates, and compare it to the PMI under the full model to approximate the percent increase in PMI achieved by using the full behavioral phenotype. We find that under the best-fitting models, BS6 and BS8, the estimated gains in PMI are 25% (2%, 3%, and 3% SEs respectively) (Figure 3.1c). The models with fewer behavioral strategies, which also have significantly lower predictive power, show lower increases in PMI.

What is the source of the additional information uncovered by modeling social interactions using combinations of behaviors? One implication of the behavioral strategy model is that there are distinct routes by which animals can arrive at similar rates of a behavior. The rate at which an animal engages in passive contact is predicted independently by the utilization of two different behavioral strategies: S2, a frequent strategy in which affiliation is relatively rare; and S4, a less common strategy of frequent interaction with lower-ranked animals. In the population we see that high rates of passive contact are associated with high use of either S2 or S4, or modest utilization of both (Figure 3.3a). We see similar patterns of results in other behaviors, such as feeding, which is predicted independently by high use of the two most risky foraging strategies, S1 and S5 (Figure 3.3b). These results indicate that the behavioral strategy model is disentangling distinct types of behavior that are not distinguishable when a behavior is considered in isolation.

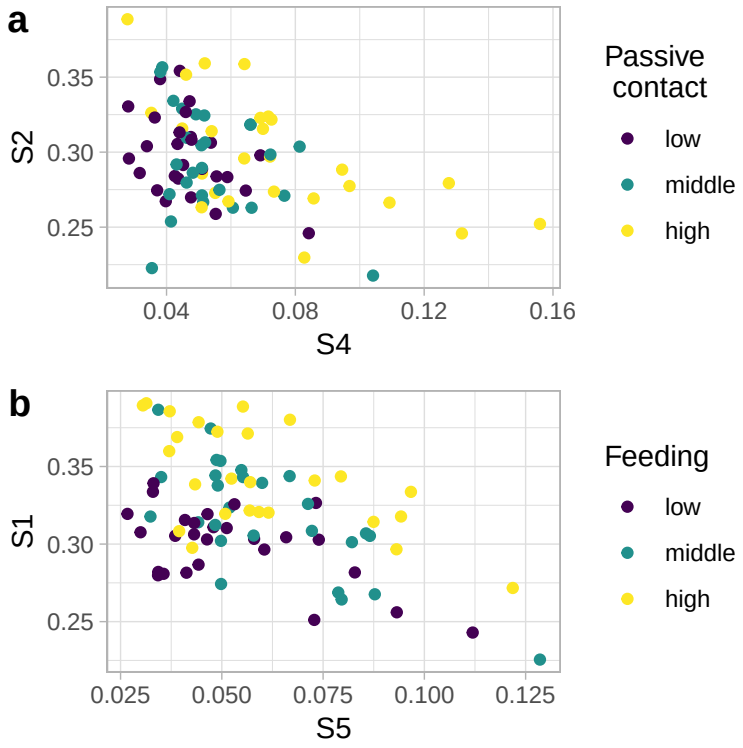

Figure 3.3: Posterior mean use rates for strategies from model BS6 across the population, with color indicating the animal's rates of the behaviors a) passive contact, and b) feeding. Rates of Passive Contact are discretized into tertiles.

Alternative model for combinations of behavior

In previous analyses we modeled the use of combinations of behavior in social interactions in terms of a small number of strategies which are shared across the population. However, this is only one possible method for modeling the way in which behaviors co-occur in combinations, and it may be that the results described above are an artifact of our particular choice of model. Accordingly, we also developed an alternative model in which the relationships between individual behaviors are modeled in terms of their pairwise correlations. This correlative model attempts to estimate all second-order interactions between behaviors, as opposed to BSK which estimates a small number of higher-order interactions between many behaviors at once.

We found that the BS with six or more strategies have significantly better predictive performance than the correlative model (minimum WAIC difference 244.9, SE 33.7), suggesting that BSK is a more appropriate model for rhesus macaque behavior. Nonetheless, we find similar results between the correlative and BSK models with regards to the impact of modeling combinations of behavior. As with BSK, we compare the full correlative model to a reduced model in which the second-order interactions between behaviors are assumed to be the same across all animals. We find that the full model has significantly better predictive power than the reduced model (WAIC difference 65.6, 16.7 SE). We also estimate PMI in the correlative model, and find that including combinations of behaviors in behavioral phenotypes increases the information gained by 25% (3% SE).

Discussion

In this chapter we developed a methodological framework for quantifying the social phenotypes of non-human primates in terms of the combinations of behaviors that characterize social interactions. We find that a model using six behavioral strategies, defined as a set of probabilities of individual behaviors, is able to capture patterns of behavior interpretable in terms of strategies of varying riskiness for foraging and socializing. Individual variability

in the use of strategies was not reducible to variability in rates of behaviors considered in isolation. By modeling the use of combinations of behavior, we increase the amount of information regarding phenotypic differences between animals by approximately 25%. These results suggest the utility of the methods developed here for understanding biological correlates of social behavior by providing greater behavioral diversity from which associations can be estimated, and by mapping biological correlates to the fine-grained details of social interactions.

To render tractable the analysis of combinations of behavior, we relied on a parametric model which necessarily imposes restrictions on the structure of animal interactions. The use of such models is currently unavoidable due to the logistical difficulties of collecting large amounts of data from free-ranging animal populations and the breadth of the primate social repertoire, but it also leads to the question of to what extent the results obtained depend on the specific choice of model. We find here that two classes of models which make very different assumptions regarding the structure of behavior produce very similar results, with both finding strong evidence that animals differ in the ways that they combine behaviors and estimating the amount of information gained about individual animals from considering combinations of behavior at 25%. These convergent results suggest the observed increase in PMI is not artifactual, but we possess no guarantee that some other method not considered here may provide a different result. It may be possible to obtain a non-parametric estimate of PMI in smaller, captive populations of animals with more restrictive environments, where each individual animal can be studied intensively. For example, Koski (2011) studied the behavior of several captive chimpanzees living in different zoos, and for one population had collected an average of 70 hours of focal observation compared to an average of 20 hours of focal observation per subject in the current study. This amount of data is equivalent to 420 ten-minute focal samples per animal, with which it may be feasible to independently estimate the rates of all combinations of six behaviors (or equivalently, 64 unique combinations of behavior). Such an analysis could then assess the bias associated with the parametric methods used here. However, any method that does not rely on parametric assumptions

on the structure of social interactions would still need to protect against overfitting, as we discuss below.

Though the primary goal of the methods developed here are to identify ways in which animals differ in their social interactions, it is important to note that identifying such differences in phenotypes is not in itself an indication that a method has been successful. A method can maximize the inferred differences between animals by overfitting the observed behavior of each animal, thus producing apparent differences in phenotypes that cannot be replicated between data sets. It is therefore important to protect against overfitting before evaluating the phenotypic differences obtained from a given model. We adopt two techniques to protect against overfitting: first, each model we considered features regularization in the form of hierarchical priors which pool the estimated phenotypes towards to population average (Gelman & Hill, 2007; Gelman et al., 2012), and second, we quantitatively compare models regarding their predictive power while correcting for model complexity, i.e. the capacity of the model to overfit (Watanabe, 2010). As a general rule, inferred differences between animals should only be regarded as plausible if the method by which they were estimated involves at least one, and ideally several, protections against overfitting. In the case where different methods are similar predictive power (or some other metric for model comparison) but give differing estimates of the amount by which phenotypes differ, then a single estimate could be produced using a method for multimodel inference such as Bayesian model averaging (Burnham, Anderson, & Huyvaert, 2011; Kass & Raftery, 1995). We did not need to use such methods here, because even though the behavioral strategy models with six or more strategies had similar WAIC scores they also produced very similar estimates of PMI.

Although here we focus on social interactions in the unconstrained setting of free-ranging animal populations, the logic of the models and analyses developed here also apply to social interactions studied in laboratory contexts. Historically, laboratory studies of social interactions have focused on a small number of predefined behaviors of interest, such as in the study of aggression and mating in rodents (Kruk, 1991; Hashikawa, Hashikawa,

Falkner, & Lin, 2016; Nelson & Trainor, 2007). Popular paradigms in this domain, such as the resident-intruder and standard opponent tests, involve placing two animals into a common environment for a short period of time and observing their interactions (Crawley, 2007). Such interactions are directly amenable to analysis under the methods proposed here, which could permit comparison across species regarding the types of strategies they deploy in social interactions and to what extent animals vary in such strategies. Recently, researchers have also begun investing the neural basis of social interactions among rhesus macaques in laboratory settings (Chang et al., 2015; Iqbal et al., 2019), as well as examining behavior in settings where animals can move about their environment in increasingly less constrained manner (Yin, Tseng, Rajangam, Lebedev, & Nicolelis, 2018). Applying the methods developed here to interactions among laboratory rhesus macaques would help bridge the gap between laboratory and field studies of social interactions.

Finally, in order to rigorously test and quantify our ability to uncover novel phenotypes from combinations of behavior in social interactions, here we restricted ourselves to examining a relatively small number of macaques from the same social group and of similar rank, a relatively small number of behaviors, and treated the occurrence of behaviors within focal observations as binaries. In the following chapter, we will show how the analysis of behavioral strategies can be extended to relax these restrictions and study large numbers of behaviors among hetergeneous populations of animals, as well as estimate the effects of biological and environmental covariates on social interactions.

CHAPTER 4 : USING MACHINE LEARNING TO DISCOVER LATENT SOCIAL PHENOTYPES IN FREE-RANGING MACAQUES

Abstract

Investigating the biological bases of social phenotypes is challenging because social behavior is both high-dimensional and richly structured, and biological factors are more likely to influence complex patterns of behavior rather than any single behavior in isolation. The space of all possible patterns of interactions among behaviors is too large to investigate using conventional statistical methods. In order to quantitatively define social phenotypes from natural behavior, we developed a machine learning model to identify and measure patterns of behavior in naturalistic observational data, as well as their relationships to biological, environmental, and demographic sources of variation. We applied this model to extensive observations of natural behavior in free-ranging rhesus macaques, and identified behavioral states that appeared to capture periods of social isolation, competition over food, conflicts among groups, and affiliative coexistence. Phenotypes, represented as the rate of being in each state for a particular animal, were strongly and broadly influenced by dominance rank, sex, and social group membership. We also identified two states for which variation in rates had a substantial genetic component. We discuss how this model can be extended to identify the contributions to social phenotypes of particular genetic pathways.

Introduction

Understanding how biological and environmental factors shape human social phenotypes – the ways in which we interact with others and in which others shape our own behavior – is a topic of intense interest in neuroscience. While neuroscientific research has made impressive strides in identifying genetic pathways and brain areas that contribute to specific social phenotypes, traditional approaches are limited in their ability to assess how biological and environmental factors contribute to naturalistic social behaviors in real-world environments. Animals used in laboratory research may differ from humans in the brain areas associated

with social cognition, and live in laboratory environments with much less varied and dynamic social interactions than natural environments (Morgan et al., 2007). Research in humans faces similar limitations, as comprehensive, in-depth observations of human social behavior in natural environments is both logistically and ethically problematic, while controlled laboratory tasks take place in restricted and low-stakes environments which may leave the true social environment unobserved.

An alternative approach to understanding biological contributions to human social behavior is to study free-ranging animals whose biology and social behavior more closely approximate our own. Free-ranging non-human primates (NHPs) provide naturalistic and human-relevant variability in social behavior that cannot be modeled in the laboratory or easily (or ethically) collected in humans. With free-ranging NHPs, it is possible to collect comprehensive data on many different aspects of social behavior, as well as both noninvasive and invasive biological samples and life history information that can be used to disentangle genetic and environmental influences. However, natural social behavior is both noisy and high-dimensional, presenting severe challenges for traditional methods of data analysis. Standard methods for relating genetic variation to traits take an atomistic view by measuring genetic influences on each measured variable individually. While this approach makes sense for specific, well-defined phenotypes of interest such as educational attainment (Rietveld et al., 2013) or height (Lango Allen et al., 2010), it is less useful for high-dimensional data where any individual dimension is of limited or unknown significance.

In the case of natural behavior, continuous streams of activity are necessarily segmented into many discrete behavioral categories during measurement, in many cases with no specific behavior being of special interest relative to the whole. The challenge is thus to develop a rigorous and biologically meaningful approach for reconstructing a "whole" behavioral state based on discrete, atomized behaviors. Such a challenge can be viewed as a type of latent structure learning, in that there is some underlying structure that is not directly observed but that ties together many different individual pieces of data. Models of this

sort are increasingly popular in the field of probabilistic machine learning (Murphy, 2012). Though machine learning models are most frequently used for prediction, here we adapt methods from the machine learning literature to infer a meaningful and useful structure in observational data sets.

Any source of genetic variation is unlikely to affect only a single discrete social behavior. A genetic variant associated with grooming behavior, for instance, may also be associated with increased time spent in the proximity of other animals in general. Genetic influences are more likely distributed across many different individual social behaviors than localized to specific types of actions (Brent et al., 2014). Furthermore, in the case of observations of natural social behavior, it is important to recognize that each type of behavior does not occur in isolation. Rather, the set of behaviors that occur during a given observational session may reflect the current situation or behavioral state the animal is in. The relationship between any given behavior and an underlying social phenotype thus may vary with context. For example, adjusting a watchband while speaking to a group of people may be a nervous tick reflecting social anxiety, while adjusting a watchband in one's bedroom might reflect only an ill-fitting watchband.

Thus, it is critical for any useful model of observed social behavior to reflect the complex interrelatedness of individual behaviors, while retaining identifiability and scalability to large data sets with dozens of individual behaviors and thousands of observations. To achieve this goal, we here adapt a family of models from machine learning known as topic models. Topic models are tools for describing the topics that are covered by a corpus of documents and to what degree each document is focused on each topic (Blei, Ng, & Jordan, 2003; Blei, 2012). In our analysis, each animal is analogous to a document, while the animal's social phenotype, or the weights assigned to different behavioral states, correspond to the weights a document assigns to each of the different topics. Under a topic model, each word in a document belongs to a topic, just as in the current model each observational session belongs to a behavioral state. Of the many topic modeling variants that exist in the literature, the

current model is most closely related to the logistic normal topic model (Lafferty & Blei, 2006; Chen et al., 2013). Our model differs from the standard logistic normal topic model in that we do not explicitly model correlations between topic weights (here, weights on behavioral states); instead, our model uses a hierarchical regression layer to incorporate outside information about documents (here, animals) in predicting their phenotypes. To the best of our knowledge, no previous topic model has included hierarchical regression at the level of topic weights (but see Rodríguez and Dunson (2011) for a related model).

Recent research on behavioral phenotypes in NHPs has used principle components analysis (PCA) to identify latent structure in natural behavior (Seyfarth et al., 2012; Brent et al., 2014), as well as in the study of animal behavior generally (Budaev, 2010). However, that research relied on rates of behaviors averaged across many different observational periods on the same animal. This means that the relationships among behaviors and the patterns in which they co-occur within a single observation are lost. That is, PCA captures variation among animals, but at the cost of losing all information on variability at the level of individual observational periods. By contrast, the hierarchical nature of topic models allows us to explicitly capture variability at both the level of individual animals and the level of observational sessions, in the form of the phenotypes used to represent individual animals and the behavioral states used to characterize observations, respectively.

Our behavioral data is a large set of animal observations taken from the free-ranging rhesus macaque (*Macaca mulatta*) colony on Cayo Santiago Island off the coast of Puerto Rico. Rhesus macaques provide an excellent model species for studying biological and environmental influences on human social behavior due to their extensive use in lab and field research; neural circuitry that is homologous to that of humans; and complex social behaviors that are critical to their biological success (Brent et al., 2013; Widdig et al., 2004; Carpenter, 1942; Bernstein & Sharpe, 1966). Because demographic, pedigree, and genetic data are also available for the colony (Brent, 2010), this detailed record of natural social behavior provides a unique opportunity for modeling genetic, demographic, and environmental influences on

complex social phenotypes.

Materials and methods

Study site and subjects

The studied population is a colony of rhesus macaques (*Macaca mulatta*) living on the island of Cayo Santiago, a 37-acre island 1 km off the southeastern coast of Puerto Rico. This is a free-ranging, freely-breeding population with known pedigrees, rich data on life histories and fitness, and extensive genetic and observational data on behavior. The colony was founded in 1938 with a population of 409 Indian-origin rhesus macaques and is currently maintained by the University of Puerto Rico Medical Sciences Campus (Rawlins & Kessler, 1986). The current population numbers approximately 1600 animals self-organized into six different social groups. Approximately 600 of the animals are adults of age six or above, and 600 are juveniles between the ages of one and five. Researcher and caretaker intervention in the population is minimal. Animals in the colony are provided commercially available monkey chow daily and unlimited access to water. Animals are handled only during designated annual trapping periods, during which infants are tagged for identification and population control can be implemented via the removal of small numbers of animals.

The data used in this study were 8205 ten-minute focal observations collected from adult (age $>$ 6 years) male and female macaques from three social groups (F, HH, and R). Observational data was collected from group F during 2012 and 2013 and from group HH and R during years 2014 and 2015, respectively. A total of 227 macaques (71 male) were represented in this data set, with 106 from group F (41 male), 41 from group HH (11 male), and 80 from group R (19 male). The number of focal observations per animal ranged from 80 (approximately 13.33 hours) to 11 (approximately 1.83 hours), with an average of 36.15 focal observations per animal (approximately 6 hours).

Observational data

The data is comprised of ten-minute focal samples, wherein a trained researcher tracks an individual monkey (the focal animal) for a ten-minute period while comprehensively recording the animal's behaviors (Brent, 2010; Altmann, 1974). Once each ten-minute session is complete, the observer moves on to a new focal animal. Observers recorded the times at which the monkey engaged in any behaviors specified by a predetermined ethogram, as well as the identities of nearby animals during the focal observation (Brent, 2010). The order in which animals were observed was semi-randomized to equalize the times of day and year that animals were observed.

This research complied with protocols approved by the Institutional Animal Care and Use Committee of the University of Puerto Rico (protocol #A6850108) and adhered to the legal requirements of the United States of America.

Behaviors

The ethogram used to guide data collection consisted of both social and self-directed behaviors. Social behaviors were further classified into affiliative and agonistic behaviors. Each social behavior was further divided based on whether the behavior was initiated by another animal and directed towards the focal animal, or initiated by the focal animal towards another animal. A total of 30 behaviors were used in the present study, including behaviors divided by giving versus receiving. Only social behaviors involving another adult macaque were used in this study; behaviors directed at infant or juvenile macaques or human researchers were not included. The behaviors were defined as follows:

- Self-directed behaviors
  - Scratch: Rapid and repeated movement of the nails of a hand or foot across the skin.
  - Self groom: Running of hands or mouth through one's own hair for at least 5 s.
  - Feed: Searching for, manipulating, holding and ingesting food items, including

water, for at least 5 s.

  - Travel: Movement from one location to another over a distance of at least 5 m.

- Affiliative behaviors

  - Approach: One individual approaches another to within arms' reach (2 m) without physical contact, and remains within that distance for at least 5 s.

  - Leave: Exiting a 2 m area around another without an agonistic interaction.

  - Affiliative vocalization (AffilVocal in figures): Emiting a friendly vocalization in the form of either a grunt, girney, vocal exchange, or lipsmack. Individuals will often emit many vocalizations in short succession.

  - Groom: Running the hands or mouth through the hair of another monkey for at least 5 s.

  - Passive contact: Sitting or lying in physical contact with another animal without grooming.

  - Social proximity: Number of time points out of three at which the focal was observed to be within 2 m of at least one other animal. The time points were at the beginning, middle, and end of the focal observation.

  - Proximity group size: Number of unique animals with whom the focal animal shared social proximity as defined above.

- Agonistic behaviors

  - Threat: One individual threatens another with one or a combination of staring, barks, head bobs, and opening one's mouth with covered teeth.

  - Avoid: Moving out of the way of another before they come within 2 m.

  - Displacement: Similar to avoid, but within 2 m.

  - Fear grimace: Submissive facial expression wherein lips are retracted horizontally to expose teeth.

  - Submit: Leaning away from another or crouching while raising hindquarters towards another.

  - Noncontact aggression: A lunge, charge, or chase that does not result in direct physical contact.

  - Contact aggression: Direct physical contact such as a bite, hit, push, or grab.

For more information on the behaviors defined above, see Brent, 2010.

Genetic data

We obtained animal pedigrees from a long-term database maintained by the Caribbean Primate Research Center (CPRC). From the founding of the population up through 1992, maternal identity was ascertained by behavioral observations, such as nurturing behaviors and lactation. For most macaques born after 1992, both maternal and paternal identity were ascertained genetically through the analysis of 29 microsatellite markers. Previous studies revealed a 97.4% agreement rate between maternity determined by behavior and by genetics (Brent et al., 2014). The population pedigree was used to generate a kinship matrix for the animals in the present study via the R package kinship2 (Sinnwell, Therneau, & Schaid, 2014). The kinship matrix was then element-wise multiplied by two to create the genetic covariance matrix used to estimate the heritability of behavioral states.

A model for social phenotypes

At the heart of the model is a finite set of latent behavioral states that determines the rates at which different behaviors occur. An animal's social phenotype is represented in the model as the rate at which that animal experiences the different behavioral states. We describe each component of the model in detail below. In this section, we first present a conceptual description of the model, and the description of the mathematical implementation follows in section 2.6.

Each ten-minute focal sample is assumed to belong to one and only one behavioral state; that is, during a given focal observation, the focal animal is assumed to be in a particular behavioral state. Each state is defined by a set of parameters that determines the probabilities that each of the discrete behaviors will occur, and, should they occur, how frequently or in what amount they will occur during a focal observation. In the data analyzed here, we used a categorical distribution over various levels of behavioral quantities, but the algorithm can trivially be modified to use different data likelihoods, such as a Poisson distribution for data in count form, a normal distribution for continuous data, and so on. For details on the

coding of discrete behaviors in the current study, see section 2.5 below.

Every animal has a different probability distribution over behavioral states, which describes how often it finds itself in each of those states. This probability is what we call the animal's social phenotype. An intuitive interpretation of this aspect of the model is that at the beginning of a focal observation, the focal animal flips a K-sided die where K is the number of states in the model. Whichever side the die lands on determines which behavioral state the animal will be in for the duration of the focal observation. While the K behavioral states are common across all animals, each animal has a unique die with different probabilities of landing on any given side compared to any other animal.

Ultimately, we are interested in not just describing social phenotypes and behavioral states, but also understanding how they are associated with other variables of scientific interest, be they morphological, demographic or genetic. Accordingly, our model allows the phenotypes of each animal to be influenced by covariates. We accomplish this by having the probabilities in the animals' phenotype be determined by a mixed-effects multinomial logistic regression. The regression model incorporates the influence of covariates as fixed effects. Two random effects components are also included in the model. The first describes the intercepts of each of the K states, which determine how typical each state is across the population as a whole. The second describes animal-specific error terms for each state, which capture how each animal deviates from the predicted phenotypes based on its covariates and the population average. This mixed effect formulation allows the model to refine the predicted phenotypes for each individual animal by pooling information across the entire population, as represented in the data (Gelman & Hill, 2007).

In addition to the random effects terms described above, the model can also incorporate a third random effects component based on groupings or relationships among animals. Here we use this term to incorporate genetic effects into the model so that the heritability of phenotypes can be calculated as in the popular "animal model" in behavioral ecology (Kruuk, 2004).

A major constraint in the model is that during a ten-minute focal sample, the focal animal is confined to only one behavioral state. This constraint is equivalent to assuming that the rate of switching between behavioral states is low enough that the probability of changing states during a consecutive ten-minute period is effectively zero. Though this constraint is not particularly realistic, attempting to infer state shifts within focal samples would dramatically increase the computational complexity of the model. Furthermore, because many individual behaviors are quite sparse—occurring on average at a rate of less than once per focal observation—state transitions within focal samples are likely to be very difficult to detect. We therefore justify this simplifying assumption for the computational efficiency and scalability it permits, though it can be relaxed in the future or with different data.

Data processing and likelihood

To construct the input to the model, we calculated the total amount of each behavior present in each observation, and converted those amounts into ordered levels. The exact procedure is described below:

1. Construct a data matrix with a row for each focal observation, and a column for each behavior in the ethogram.
2. For each focal observation:
    (a) For each "event" behavior, count the number of times that behavior occurred during the observation.
    (b) For each "activity" behavior, calculate the total proportion of the focal observation spent engaged in that behavior.
    (c) Populate the associated row in the data matrix with these values.
3. For each behavior:
    (a) Calculate quintiles, e.g., 20th percentile, 40th percentile, etc., of the values in that behavior's associated column in the data matrix.
    (b) Also calculate the 1st and 99th percentile of the behavior to make high and low outliers.
    (c) Bin focal observation using the quantiles calculated above as cutpoints, e.g., values $\leq$ 1st percentile being 1, $>$ 1st and $\leq$ 20th percentiles being 2, etc.

This procedure divides each behavior into up to seven levels, with 1 being the smallest amount of a behavior and 7 being the most. However, in practice many behaviors occurred so infrequently that the bottom 50% or more of the data were all zeros. This meant that 19 of 30 behaviors were split into three levels representing 0, 1, and ¿1 occurrences, and no behavior was divided into more than 6 levels.

A behavioral state consists of a set of categorical distributions, one for each behavior. Each categorical distribution has as many parameters as the behavior has levels that determine the probability of each level of the behavior occurring under that state. This allows each behavioral state to be extremely flexible; a state need not specify a narrow range of values or a specific distributional shape for every behavior. A state may be associated with a very specific amount of one behavior, but also consistent with a wide range of amounts of a different behavior.

The categorical distribution we used as the data likelihood can be easily changed to any distribution with a conjugate prior. For example, one might use the Poisson distribution for event behaviors and the zero-truncated normal distribution for activity behaviors. We chose the categorical distribution here because, as can be seen in Figure S1, many behaviors were zero-inflated with long right tails, which cannot be captured by Poisson and normal distributions.

Mathematical description of the model

The content of a focal observation is represented by $y^{(i,f)}$, a vector of length B, where B is the number of individual behaviors considered in the model. Element $y_b^{(i,f)}$ is the amount of behavior b that occurred in that focal observation. Each focal observation in the data belongs to a single behavioral state. Formally, this means that each observation f of animal i is associated with a latent variable $z_{i,f}$ which can take on values 1 through K, where K is the number of behavioral states in the model. This value denotes which of the K behavioral states to which the focal observation belongs. Given a value for behavioral state $z_{i,f}$ we

can write the data likelihood of the focal observation using the parameters defining the behavioral state:

$$\mathrm{p}(y^{(i,f)}|z_{i,f}=k,\theta^{(k)}) = \prod_{b\in 1:B} \theta^{(k,b)}_{y_b^{(i,f)}}\,, \tag{4.1}$$

where $\theta^{(k)}$ is the set of all parameters associated with state k, $\theta^{(k,b)}$ is the vector of probabilities for different levels of behavior b in state k, and $\theta_l^{(k,b)}$ is the probability that behavior b occurs at level l in state k. Note that this data likelihood is simply a product of categorical likelihoods, one for each behavior. Because each focal observation is statistically independent conditional on the state assignments, we can write the complete likelihood for animal i as simply the product of the likelihoods of the focal observations:

$$\mathrm{p}(y^{(i)}|z_i,\theta) = \prod_{f\in 1:n_i} \mathrm{p}(y^{(i,f)}|z_{i,f},\theta^{(z_{i,f})})\,, \tag{4.2}$$

where n_i is the number of focal observations for animal i.

We now turn our attention from the likelihood to the phenotype, which is the prior probability of a focal observation f of animal i belonging to each of the k behavioral states. We can write this prior probability as

$$\begin{aligned} \mathrm{p}(z_i|\pi_i) &= \prod_{k\in 1:K} \pi_{i,k}^{n_{i,k}}\,, \\ n_{i,k} &= \sum_{f\in 1:n_i} \mathbf{1}(z_{i,f}=k) \end{aligned} \tag{4.3}$$

where $\pi_{i,k}$ is the prior probability that animal i finds itself in state k and $n_{i,k}$ is the total number of focal observations of animal i belonging to state k. The probabilities $\pi_{i,k}$ are themselves determined by via multinomial logistic regression:

$$\begin{aligned}\pi_{i,k} &= \frac{\exp(\eta_{i,k})}{\sum_{k\in 1:K}\exp(\eta_{i,k})}\,,\\ \eta_{i,k} &= \alpha_k + X_i^T\beta_k + u_{i,k} + \epsilon_{i,k}\,.\end{aligned} \tag{4.4}$$

Here η_i is a vector of unbounded propensities for animal i to fall into each behavioral state, which are transformed by the softmax function into the probability distribution over states, π_i. X_i is a vector of animal-specific covariates, with the state-specific fixed effect regression coefficient vectors β_k. The quantities α, u, and ϵ are random effects terms representing baseline state propensities, genetic effects, and individual animal effects, respectively. These random effects are given normal distributions with the variances as free parameters to be estimated:

$$\begin{aligned}\epsilon_{i,k} &\sim \mathrm{N}(0,\sigma_k^2)\,,\\ \alpha_k &\sim \mathrm{N}(0,\tau\sigma_k^2)\,,\\ u_{\cdot,k} &\sim \mathrm{N}(0,\gamma_k\sigma_k^2 A)\,.\end{aligned} \tag{4.5}$$

The parameters σ_k^2 determine how much variability exists across animals in the propensity for state k (outside of variability accounted for by the covariates X_i). The parameter τ controls the extent to which the states themselves vary in their average propensities across animals. That is to say, in a model where the σ_k^2 are large and τ is small, no state will consistently have high or low probability across the population, whereas when σ_k^2 are small and τ is large, states will have similar probabilities across animals but some states will be consistently high probability and others low. Finally, the covariance matrix A is the relatedness matrix of the animals in the study population. The parameters γ_k determine how much variation in each state's propensities is accounted for by genetic effects. Further, note that the variance component parameters τ and γ_k are being scaled by the "global" variances σ_k^2. This causes the linear regression in Equation 4.4 to be fully conjugated, such that the sampler can be "partially collapse" during inference by integrating out the components of the linear regression (van Dyk & Park, 2008). See (Park & Min, 2014) for an explanation of partially collapsed samples in mixed effects regression.

As we are using a fully Bayesian approach, we must specify priors for the free parameters. We use standard conjugate uninformative or weakly informative priors in all cases:

$$\begin{aligned} \theta^{(k,b)} &\sim \text{Dirichlet}(1)\,, \\ \beta_k &\sim \text{N}(0,1)\,, \\ \sigma_k^2 &\sim \text{InvGamma}(0.005, 0.0005)\,, \\ \tau &\sim \text{InvGamma}(0.005, 0.0005)\,, \\ \gamma_k &\sim \text{InvGamma}(0.165, 0.0165)\,. \end{aligned} \tag{4.6}$$

See Gelman et al. (2008) for a discussion on weakly informative priors, though we do not use their exact priors. Note that the seemingly weakly informative prior on the γ_k parameters was chosen in order to achieve an uninformative (high variance) prior on heritability, which in this model is approximately $\gamma_k/(1+\gamma_k)$. An uninformative prior such as $\gamma_k \sim \text{InvGamma}(0.005, 0.0005)$ would actually place high prior probability mass on values of $\gamma_k/(1+\gamma_k)$ near 1. The chosen prior is roughly symmetric on $\gamma_k/(1+\gamma_k)$ and places much of the probability mass near both 0 and 1.

Model fitting

We fit the model using a custom Gibbs sampler implemented in the Julia technical computing language v0.5.0 (Bezanson et al., 2017). As standard logistic regression representation is non-conjugate and therefore cannot be sampled from using Gibbs, we use the fully conjugate latent variable formulation of logistic regression described in Polson et al. (2013) and previously applied to topic models in Chen et al. (2013). In the results reported below we ran two chains of 100,000 samples each, which were thinned to 1000 samples, with the first 100 of those discarded as burn-in.

Our model involves unsupervised classification with several hundred parameters, which means the posterior is likely to be highly multimodal. Gibbs sampling with naively chosen random starting points can often get stuck around suboptimal and idiosyncratic local maxima, which

leads to both poor inference and results that are unpredictable and unreproducible between runs. Tests with random starting points indicated that even in simple synthetic data sets, where observations fall into distant, non-overlapping clusters, Gibbs sampling alone very frequently yielded incorrect clustering whenever more than three or four behavioral states were used. In many applications of topic modeling, reproducibility and interpretability of model outputs are of secondary concern to pure predictive performance, but for scientific inference they are paramount.

In order to improve the quality and reliability of inference we first fit a simpler version of the model which we then used to generate starting points for the full model. Specifically, we fit a "flattened" version of the full model which is equivalent to assuming that all observations came from the same animal, thus discarding all information about differences between animals and population level covariates. This model is in effect a naive Bayes classifier, where each possible classification is analogous to a behavioral state. This flat model was fit by maximum likelihood (ML) using the Stan software package v2.14.0 (Carpenter et al., 2017). The initial points for this optimization step were generated by first taking the ML solution for a model with a single behavioral state, then adding independent Gaussian noise to each parameter to generate starting points for each behavioral state. Multiple fits with independent initializations were generated and the fit with the highest posterior density was used to initialize the Gibbs sampler. Specifically, for each focal observation, an assignment to a behavioral state was randomly drawn from the posterior distribution of behavioral state memberships under the ML fit, and this mapping from observations to behavioral state was used as the starting point for the Gibbs sampler. We found that in practice this procedure very reliably recovered the true parameters when applied to simulated data and provided consistent results using real data.

Repeatability analysis

Seventy-seven animals from social group F were observed for two consecutive years (2012 and 2013), and we used these animals to assess the stability of the phenotypes discovered by

our model. To accomplish this objective, we ran a separate model in which all animals from the previous analysis were included, but animals with dual observation years were permitted to have independent phenotypes for each year. We then examined the correlation between the posterior means of the phenotypes in 2012 and 2013. No heritability component was included in this model. Note that this version of the model was not used for any analysis other than heritability, as the artificial inflation of the number of animals might lead to unwarranted confidence in population-level inferences.

Simulations

To verify that the behavioral state model could correctly recover both state contents and individual phenotypes, we tested the model on synthetic data in series of five simulations using 5, 10, 20, 40, and 80 states. For each simulation, we simulate 200 individuals with 100 focal observations each, using an ethogram consisting of 30 behaviors. Each behavior was represented as a categorical distribution with 3 levels. States were generated by sampling from the following prior:

$$\begin{aligned} \lambda_l^{(k,b)} &\sim \mathrm{N}(0, 1.0)\,, \\ \theta_l^{(k,b)} &= \frac{\exp(\lambda_l^{(k,b)})}{\sum_{l' \in 1:3} \exp(\lambda_{l'}^{(k,b)})}\,. \end{aligned} \tag{4.7}$$

Similarly, individual phenotypes were generated by sampling from the following distribution:

$$\begin{aligned} \eta_{i,k} &\sim \mathrm{N}(\alpha_k, 0.0625)\,, \\ \alpha_k &\sim \mathrm{N}(0, 0.25)\,. \end{aligned} \tag{4.8}$$

These values are converted into individual probability distributions over states. Each individual's probabilities are then used to sample a state membership for each of that individual's 100 focal observations.

Because states have no intrinsic ordering, some method is required for associating states in the simulated data with a matching state in the model output before it is possible to

determine whether a state or phenotype has been accurately estimated. To accomplish this, we used a greedy matching algorithm with the folowing steps:

1. Pick an output state k' and calculate the posterior mean for each of its parameters, $\hat{\theta}^{(k')}$.
2. For each simulated state k that is not already matched with an output state, calculate the correlation between $\theta^{(k)}$ and $\hat{\theta}^{(k')}$.
3. Pick the simulated state with the highest correlation as the match for the output state k'.
4. Repeat for each k'.

After the matching procedure, we assessed how well the model recovered individual phenotypes by calculating for each individual the correlation between that individual's simulated probability of being in each state and the probabilities of being in the matching states of under the fitted model. We also assessed our ability to recover the true number of behavioral states in the 5-state simulation by comparing WAIC scores of the model with 5 states to models with 3, 4, 6, and 7 states fitted to the same data.

The model was fit to the simulated data using the same fitting procedure as used with the Cayo data. However, no regression coefficients or heritabilities were calculated.

Comparisons with factor analysis

We fit factor analysis models to the Cayo Santiago data set to compare with the behavioral state model presented here. As factor analysis and related models have no explicit hierarchical structure, they cannot separate observation-level relationships among behaviors from organism-level relationships; they can either examine how behaviors co-occur within observations, or how average rates of behaviors can co-occur between macaques, but not both simultaneously. Therefore we fit two models. The factor model 1 used the focal observations themselves as independent data points. For factor model 2, we calculated average rates of each behavior for each macaque and used macaques as independent data points. To set the number of factors used, we used 5-fold cross-validation to choose amongst 5, 10, 15, and 20

factors. For both versions 1 and 2 of the model, models with 10 factors or more performed similarly, so we used 10 factors for ease of comparison with the state model. We defined the phenotypes infered by the factor models as the factor scores of the individual animals. In the case of factor model 1, scores are associated with individual focal observations rather than individual animals, so the phenotypes were the average of the scores across each macaque's focal observations.

To estimate heritability of phenotypes under both models, we fit an animal model to the inferred phenotypes. These models used the same covariates as the state model.

To calculate conditional means for rates of behaviors under factor model 1, we sampled focal observations from a multivariate normal distribution with the means and covariance matrix fit by the factor model. Since the true data is discrete rather than continuous, for each of the sampled observations we rounded the behavior amounts to the nearest whole number occurring in the true data.

All factor analyses were carried out using the *factanal* function in R 3.4.0 using "regression" scores. (R Core Team, 2017). Animal models were written and fit using the Stan software package (Carpenter et al., 2017).

Assessing genetic and covariate influences on social phenotypes

In multinomial logistic regression, regression coefficients associated with a state are only interpretable relative to a baseline state (Agresti, 2002). This is due to the fact that a probability distribution over K states has only $K - 1$ free variables, as the probability of a single state is determined completely by that of the remaining states. If a baseline state is not specified, parameters will be unidentifiable with respect to the data and will be constrained only by the prior. Regression coefficients and genetic effects, with the coefficients and genetic effects of the baseline state subtracted, therefore reflect an influence on the relative probabilities of a each state occurring versus the baseline state.

A related issue is that the state probabilities, here the phenotypes of interest, are a nonlinear function of the underlying model parameters and covariates. The impact on the phenotypes of any given component of the model depends on the value of every other covariate and parameter in the model. Due to this nonlinearity, and because values depend on the baseline selected, parameter estimates themselves can be very difficult to interpret. However, an advantage of our Bayesian approach is that we can easily derive estimates and central credible intervals (CIs) for the influence of specific components of the model on the phenotypes of interest. We accomplished this by generating predicted phenotypes from posterior samples under different assumptions to assess the contributions of different components of the model. To assess the impact of a specific covariate on phenotypes, we generated predicted state probabilities at varying levels of that covariate while holding every other covariate fixed at its population average value (or, in the case of discrete covariates, the modal level). This process was repeated for every posterior sample of the model parameters, yielding a full posterior distribution of predicted phenotypes.

To assess which, if any, behavioral states were strongly influenced by genetic effects, we calculated a pseudo-h^2 metric to describe the amount of variance in state probabilities that was explained by genetic effects and covariates combined relative to the variance captured by covariates alone. Formally, for state k the total variance in state probabilities is $\text{var}(\pi_{\cdot,k})$, where $\pi_{\cdot,k}$ is the vector of state probabilities as defined above (Equation 4.4). We can also define partial variances based on state probabilities estimated when certain components of the model are omitted:

$$\begin{aligned} \hat{\pi}_{i,k}^{(B)} &= \frac{\exp(\hat{\eta}_{i,k}^{(B)})}{\sum_{k\in 1:K}\exp(\hat{\eta}_{i,k}^{(B)})}, \\ \hat{\eta}_{i,k}^{(B)} &= \alpha_k + X_i^T\beta_k, \end{aligned} \tag{4.9}$$

$$\begin{aligned} \hat{\pi}_{i,k}^{(u)} &= \frac{\exp(\hat{\eta}_{i,k}^{(u)})}{\sum_{k'\in 1:K}\exp(\hat{\eta}_{i,k'}^{(u)})}, \\ \hat{\eta}_{i,k}^{(u)} &= \alpha_k + X_i^T\beta_k + u_{i,k}. \end{aligned} \tag{4.10}$$

We then define pseudo-h^2 for behavioral state k as

$$1 - \frac{\text{var}(\pi_{\cdot,k} - \hat{\pi}^{(u)}_{\cdot,k})}{\text{var}(\pi_{\cdot,k} - \hat{\pi}^{(B)}_{\cdot,k})}. \tag{4.11}$$

This is the proportion of variance explained by adding genetic effects back into a model, out of the residual variance left over by a model with covariates alone. This is both closely related to a partial R^2, and to the traditional h^2 measure used in animal models, where the latter is commonly defined as the proportion of variance explained by genetic effects out of the residual variance after the effects of covariates are removed (Wilson et al., 2010). As with covariates, we calculated pseudo-h^2 for every posterior sample. Finally, we note that unlike standard h^2 and R^2, pseudo-h^2 can be negative, indicating that including genetic components reduces predictive accuracy.

Results

Simulation tests

We first verified that our model could accurately recover both the contents of behavioral states and phenotypes of individual animals by fitting the model to simulated data sets and comparing the model outputs to the ground truth. Figure 4.1 shows that for data sets of similar size as the Cayo Santiago data, the model accurately recovered all behavioral states so long as the number of states did not exceed 20. In each simulated data set with up to 20 states, every simulated state had a unique state in the fit model for which their state parameters $\theta^{(k)}$ had a correlation coefficient above 0.95. When the number of states exceeded 20, the majority of simulated states had a close match in the model's estimated states, but several simulated states had no match with correlations above 0.5. This suggests that the model was unable to find some behavioral states.

Similarly, the ability of the model to reliably recover phenotypes declined as the number of states increased. With 5 and 10 states, the correlation between fitted and simulated

phenotypes was above 0.75 for most individuals and above 0.5 for nearly all of them, while in the data sets with 40 and 80 states most individuals had correlations below 0.5. This is not surprising, as larger numbers of states means more parameters must be estimated for each individual using the same number of observations.

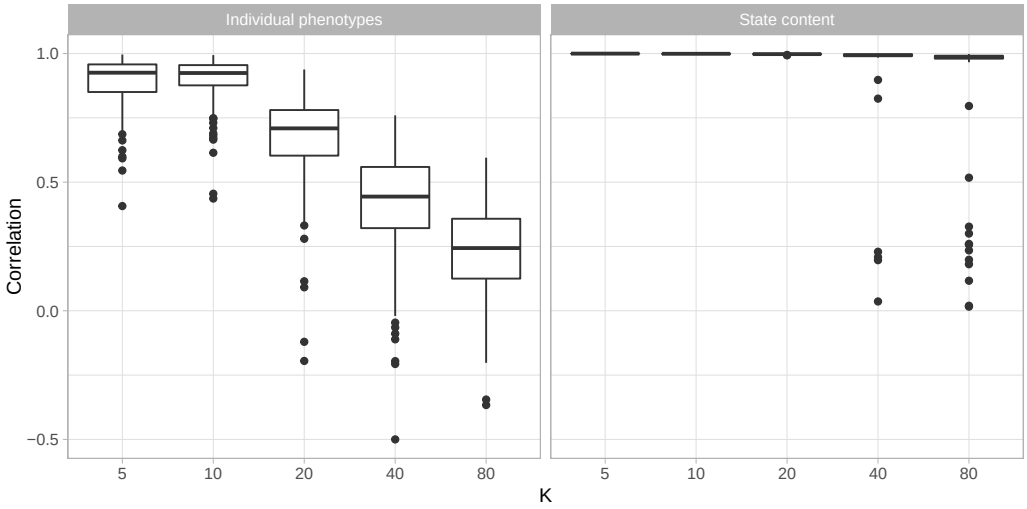

Figure 4.1: Correlations between simulated and fit phenotypes (left) and state contents (right). Both panels show boxplots, though for the state contents the values are concentrated enough that the hinges of the plots are not distinguishable. For phenotypes, each data point is an individual, while for states, each data point is a state.

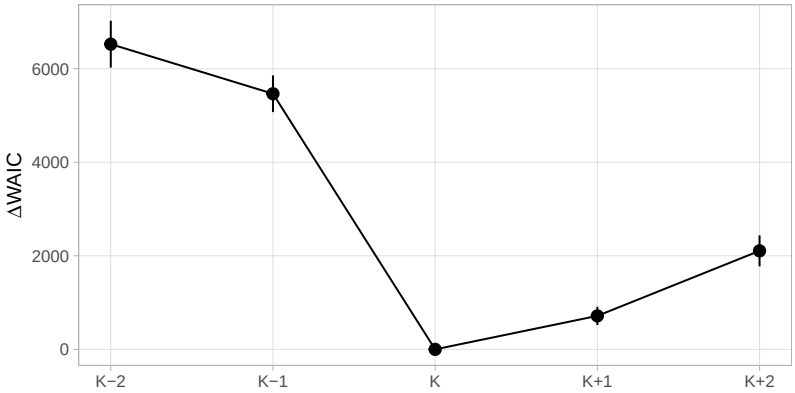

Figure 4.2: Choosing number of states for simulated data using WAIC. Y-axis shows difference in WAIC between the model with the correct number of states, K, and models with more and fewer states. Error bars represent two standard errors. Lower is better.

We also found that for the data set with five states we were able to identify the model with

five behavioral states as the optimal model using WAIC goodness-of-fit metric. See Figure 4.2 for the WAIC simulation results.

Phenotype distributions and behavioral state content

In order to choose an appropriate number of behavioral states, K, we calculated the widely-applicable information criterion (WAIC) (Vehtari et al., 2017) for models (without a genetic component) with 5, 10, 15, and 20 states, and used the number of states associated with the lowest WAIC. We identified the model with 10 states as having the lowest WAIC (Figure 4.3). Accordingly, we used 10 behavioral states for all subsequent analyses (referred to hereon as S1–S10). Figure 4.4 shows the distribution of phenotypes across the population, in the form of the rate at which each animal exhibited each behavioral state. We ordered the states in terms of how frequently each state occurred in the population, with S1 having the highest average rate of occurrence and S10 having the lowest. S1–S3 showed both high overall rates, with animals spending on average 55% of their focal observations in these states, as well as high variability in their rates across animals, with substantial numbers of animals spending more than 30% of their focal observations in a single one of these three states. S4–S9, on the other hand, all had median rates falling in a narrow range between 7% and 5.6%, with only one animal displaying a rate greater than 20% in any one of those states. Finally, S10 has a median rate of only 1.4%.

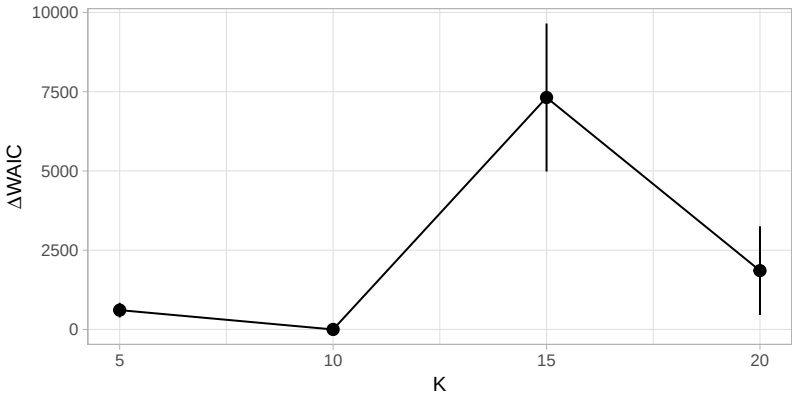

Figure 4.3: Choosing number of states for the Cayo Santiago data using WAIC. Y-axis shows difference in WAIC between the model with 10 states and models with more and fewer states. Error bars represent two standard errors. Lower is better.

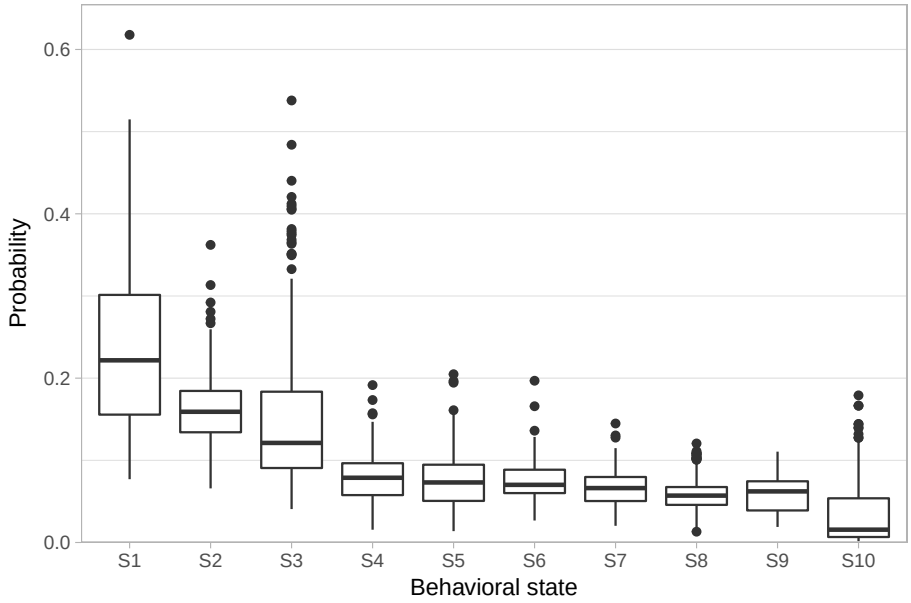

Figure 4.4: The distribution of phenotypes of the studied population. The box-and-whisker plots display the distribution of posterior mean probabilities of being in each state across all studied individuals. States are ordered by descending mean probability across the population. "Hinges" of the boxes represent 25%, 50%, and 75% quartiles.

Next we examined the contents of these behavioral states. Figure 4.5 visualizes the typical behaviors of each state in terms of their relative frequency, with very rare behaviors omitted for legibility. Such "at a glance" summaries distinguish particular features of each behavioral

state. Immediately apparent is the distinction between the more frequent S1–S3 and the infrequent S4–S9. S1–S3 paint a rather prosaic portrait of macaque life, consisting largely of self-directed behaviors, such as scratching oneself, eating, and walking. Indeed, S2 entailed doing little other than resting. The social interactions that did occur in these three states were primarily agonistic. In particular, both overt non-contact aggression and less overt agonistic actions, such as fear grimaces and avoidances, were quite common in S1. The middle infrequent S4–S9 on the other hand, displayed mixtures in varying proportions of agonistic and affiliative actions. Incidental affiliative behaviors such as approaching other animals occurred commonly throughout all of these infrequent states, while grooming, a more significant sign of affiliation, appeared concentrated in S4, S5, and S8.

Repeatability

As our goal is to identify phenotypes with underlying biological bases, it is important to show that phenotypes produced from the model are consistent within animals and relatively stable across time. To that end we compared estimated phenotypes from animals in group F from years 2012 and 2013. Phenotypes in 2012 were strongly correlated with phenotypes in 2013 for all behavioral states, with the lowest correlation coefficient being 0.64 for S1, and the largest being 0.88 for S8 ($p < 0.001$ for both).

Group, rank, and sex effects on social phenotypes

While describing the content of behavioral states is useful, scientists are often more interested in identifying sources of behavioral variability. In rhesus macaques and other NHPs, it is important to understand how much variability in social phenotypes across animals can be predicted on the basis of demographic covariates, and how much is idiosyncratic to each individual. In the present model we included age, sex, dominance rank, social group, and age-by-sex and rank-by-sex as population-level predictors of phenotypes. Dominance rank was represented on an ordered categorical scale: low-ranking animals outranked less than 50% of their social group, medium-ranking animals outranked between 50% and 80%; and

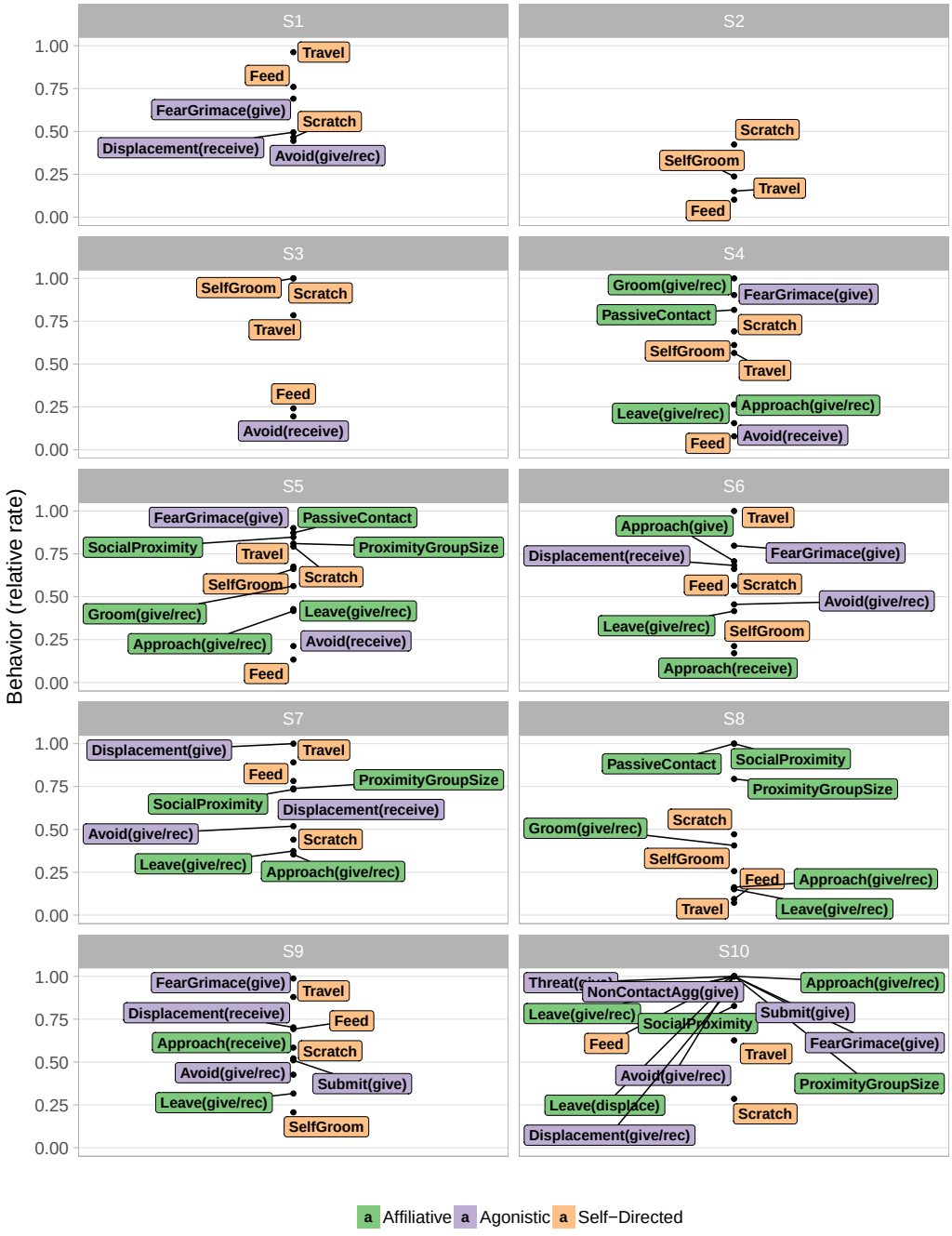

Figure 4.5: The content of behavioral states fit by the model. Relative rates for each behavior are calculated by dividing the posterior mean rates across states by the mean rate of the highest state, such that 1 represents the highest mean rate across states. For visual clarity, behaviors with relative rates below 0.05 are omitted, and behaviors for which the difference between the "give" and "receive" variants is less than 0.33 are concatenated into a single label (give/rec).

high-ranking animals outranked greater than 80%. Ranks were available annually, so for animals with multiple years of observations that changed ranks, their average rank was used. Figure 4.6 displays the regression coefficients for each covariate's influence on the probability of being in each behavioral state. Here we choose S2 as baseline as it provides a neutral default state with little in the way of positive or negative social interactions.

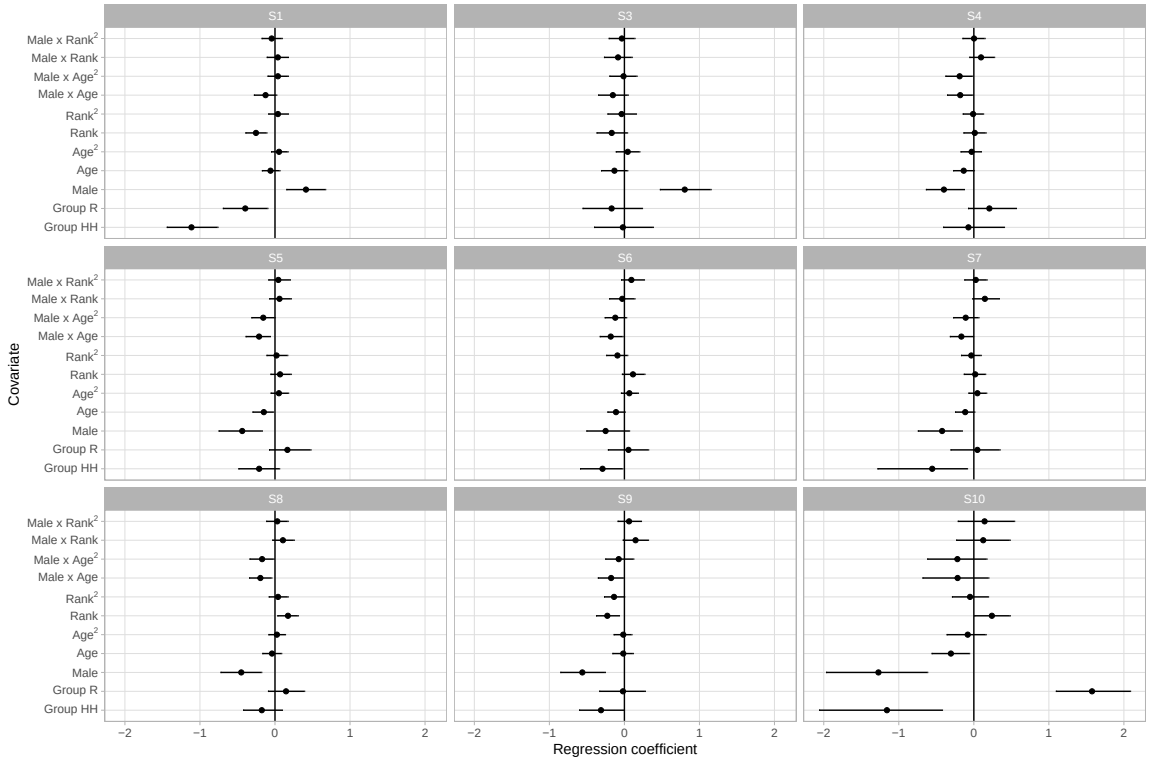

Figure 4.6: Effects of covariates on phenotypes. Regression coefficients with S2 as a baseline (see Methods). Points represent posterior means and lines, 95% central credible intervals.

The posterior distributions of the regression coefficients indicate that social group membership and sex had large influences on social phenotypes. Animals in groups R and HH had, for instance, lower relative rates of being in S1 versus S2 than animals in the largest social group, F, with regression coefficients of -0.52 ($[-0.78, -0.27]$ 95% CI) and -0.88 ($[-1.20, -0.58]$ 95% CI), respectively. As S1 had much higher rates of agonistic behavior than the unsocial S2 and typically involved very little affiliative behavior, this suggests that animals in group F were more likely to become involved in agonistic interactions. Sex

and linear dominance rank were the most common predictors of relative behavioral state occurrence rates, with four and five of the nine non-baseline states respectively showing significant effects. These states overlap, with S1, S3, and S6 being more frequent in male macaques and in lower ranking monkeys, which we note were states with higher rates of agonistic behaviors and relatively lower rates of affiliative behaviors.

Though examining the posterior of regression coefficients predicting relative odds of states can be useful, it is often more intuitive to directly examine the influence of covariates on the absolute rates at which an animal experiences behavioral states. We used this method to examine how sex, group membership, and dominance rank influenced phenotype. Figure 4.7 displays the influence of social group membership on the social phenotype of male and female animals of median age and rank. This representation of the model demonstrates that males experienced S1 and S3 more frequently than females, at the expense of slightly lower rates of all of the less frequent, more social S4–S10. S1 entailed giving and receiving aggression while moving and feeding, while S3 consisted of self-directed grooming and scratching with some amount of agonistic actions that may have served to defuse aggression. Both featured low rates of affiliative social behavior. In contrast, the states males experienced less often contained higher relative rates of affiliative behaviors and peaceful sharing of space with conspecifics. This implies that males spent more of their time avoiding confrontations and competing for resources and less time experiencing social support.

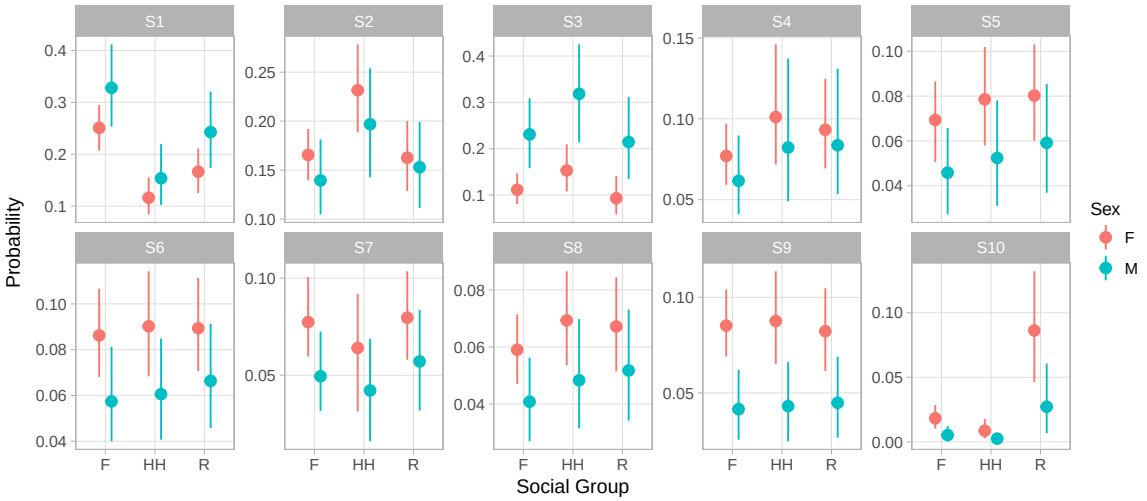

Figure 4.7: Expected probabilities of being in a state across different social groups for animals of average age and rank. Points represent posterior means and error bars, 95% central credible intervals of the expected probability.

It is worth noting that males experienced S6 at lower rates than females did, despite the fact that the S6 regression coefficient for "male" has a positive posterior mean and little posterior mass below zero (0.13 posterior mean, [−0.03,0.45] 95% CI). This illustrates the value of examining absolute state probabilities rather than regression coefficients themselves. In other words, because males spent so many focal observations in S1 and S3, the positive coefficient for males in S6 reduced the extent to which S1 and S3 crowded out S6 for males, but did not reverse the gap between sexes.

Finally, we note that S10, the least frequent of states overall, showed a dramatic difference between groups. Monkeys in group R experienced this state far more often than members of either F or HH. As before, this effect was not apparent from the raw regression coefficients, as the coefficients for HH and R membership were similar in absolute value, with posterior means of −0.91 ([−1.80,−0.14] 95% CI) and 1.45 ([0.96, 1.95] 95% CI) respectively. S10 was characterized by very high rates of feeding while maintaining close proximity to large numbers of other animals. Focal animals in S10 also engaged in high rates of agonistic behaviors such as threats and non-contact aggression, suggesting this state captured episodes

of food competition within or between groups.

Figure 4.8 shows the influence of dominance rank on the phenotype of an animal of median age in group F. Here we observed an overall pattern across behavioral states that was similar to the effect of gender, in that S1 and S3 were overall more frequent among low ranking animals than high, while S4–S10 were more frequent in high ranking animals. We also observed strong rank-by-sex interactions in S3 and S9, whereby higher rank had little effect on the rate of being in S3 for females but greatly decreased that rate for males, while the opposite was true for S9.

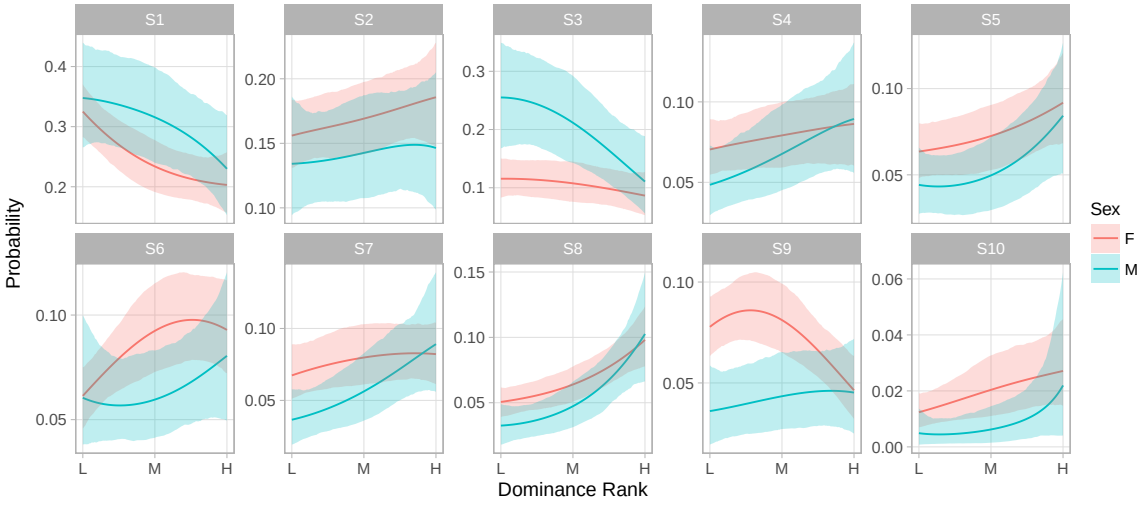

Figure 4.8: Relationship between dominance rank and the expected probability of being in a state for animals of average age in group F. Curves represent posterior means and shaded regions, 95% central credible intervals of the value of the curve at the corresponding rank.

Genetic components of social phenotypes

Finally, we sought to clarify which aspects of social phenotypes were strongly influenced by genetic factors. To quantify how much variability in the probability of being in each behavioral state could be accounted for by genetic factors, beyond the variability explained by demographic covariates alone, we calculated pseudo-h^2 as described above for each behavioral state and displayed the posterior distributions in Figure 4.9. For most states, we found very wide 95% credible intervals for the variability explained with the lower bounds

of eight of the ten states falling below 10%. For these states, we could not effectively rule out negligibly small contributions of genetics in determining behavioral state probabilities. However, S5 and S6 showed reliably high pseudo-h^2 of 0.69 ([0.43, 0.80] 95% CI) and 0.6 ([0.24, 0.76] 95% CI), respectively.

S5 was associated with high levels of affiliative behavior, involving high rates of friendly vocalizations, grooming, and passive physical contact, as well as being in close proximity to a high number of conspecifics for long periods of time (Proximity Group Size and Social Proximity, respectively, see Appendix for definitions). S5 also showed low rates of most agonistic interactions. Conversely, S6 was characterized by a more diverse selection of behaviors, entailing high rates of less confrontational forms of agonistic behavior (e.g., giving and receiving fear grimaces and displacements) and more actively confrontational behavior (e.g., giving threats and noncontact aggression, as well as submissive gestures). Nonetheless, animals in S6 were not deterred from affiliative approaches.

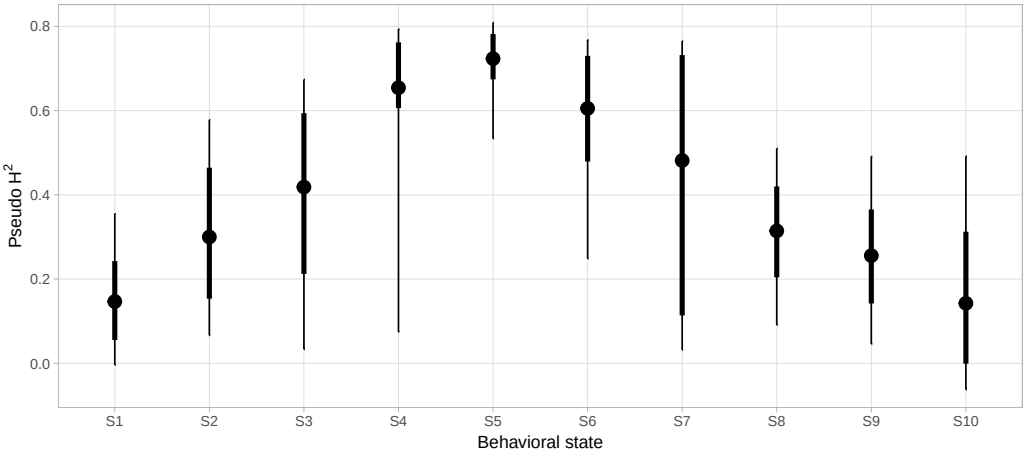

Figure 4.9: Genetic influences on the probability of being in a state in the studied population. Points represent posterior means, thick error bars, 66% central credible interval (roughly equivalent to 1 standard error), and thin error bars, 95% central credible intervals. See Methods for definition of the pseudo-h2 measure.

Comparison with factor analyis

One advantage of the state model over factor analysis and related methods is that it can capture a wider range of relationships among behaviors. Figure 4.10a gives an example of this using the behaviors Travel, Feed, and Noncontact Aggression (received). The relationship between traveling and feeding is nonlinear, with the rate of traveling increasing with the rate of feeding for low levels of feeding, but sharply decreasing at very high rates of feeding. Moreover, this relationship is moderated by noncontact aggression received. During focal observations when the animal is not feeding, animals travel much more when receiving aggression, but when feeding occurs received aggression has little impact on time spent traveling. The state model captures both the nonlinear relationship between travel and feeding as well as the interaction with noncontact aggression received, though it does notably underestimate how quickly time spent traveling decreases at high levels of feeding. Factor model 1, however, can only represent linear relationships (Roweis & Ghahramani, 1999) and so captures only the positive correlation between traveling and feeding at low levels of feeding. We did not assess factor model 2 in this way; because factor model 2 is a model of average behavior rates by individuals rather than a model of focal observations themselves, we saw no clear method for generating predictions for these observation-level relationships.

Finally, we also compared the state model to the factor models in the repeatability and heritability of the phenotypes generated by the model. As shown in Figure 4.10b, the phenotypes under the state model showed higher repeatability than those of either factor model. As shown in Figure 4.10c, the factor model yielded no phenotypes with heritability estimated at above 0.5, while the state model and factor model 1 phenotypes showed broadly similar distributions of heritabilities. However, given the high uncertainties associated with heritability estimates under all models, any conclusions about differences among models must be tentative at best.

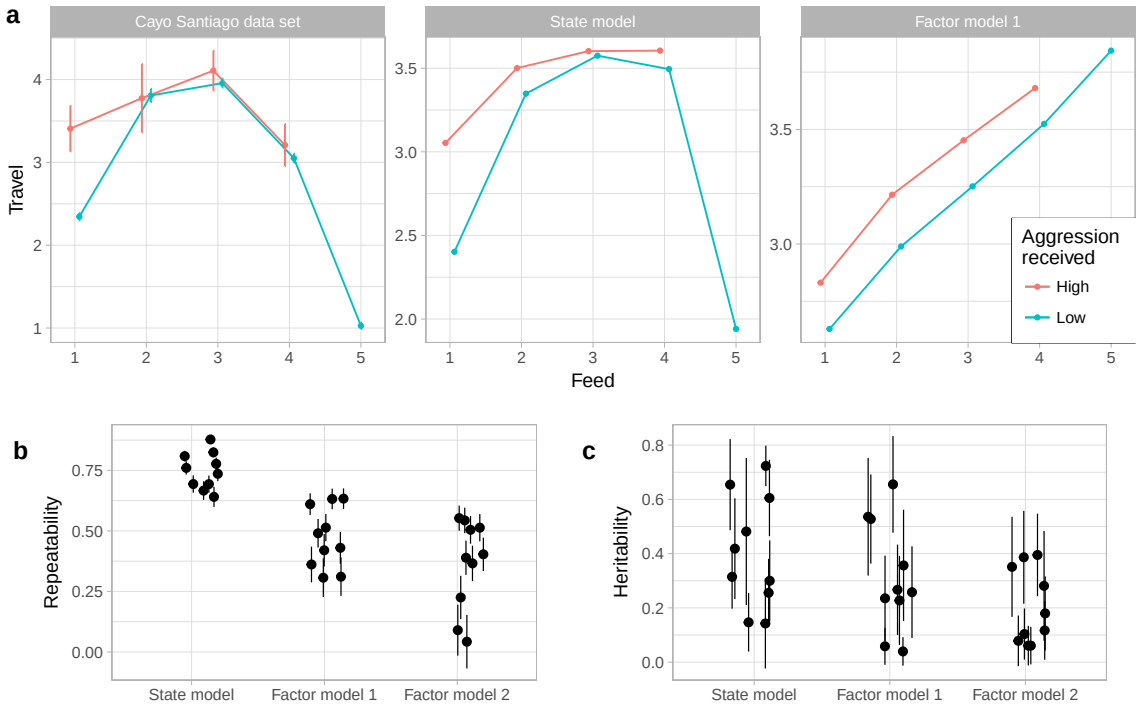

Figure 4.10: Comparisons between the state model and factor analysis models. (a) The leftmost panel shows observed means and standard errors of rates of Travel at varying levels of Feed and Noncontact Aggression (received); The right two panel shows the expected levels of Traveling under the state model and factor model 1; (b) Repeatability, as measured by the correlation between 2012 and 2013 phenotypes, for the three models; (c) Heritability estimates of phenotypes from the three models. Error bars in all panels represent one standard error.

Discussion

The influence of typical biological and environmental factors on social behavior is likely to be many-to-many, with any given factor impacting many different behaviors together rather than any specific behavior in isolation (Yang et al., 2010; Davies et al., 2011; Benjamin, Cesarini, van der Loos, et al., 2012; Neale et al., 2012). Dealing with high-dimensional, structured behavioral data will therefore be important for understanding the biological bases of behavior. The field of primatology has decades of experience collecting detailed, high-dimensional data on social behavior in species with complex, human-like social behaviors, and

this data may prove invaluable for linking fundamental biology to complex social behavior. However, extracting useful structure from natural, high-dimensional and relatively noisy behavioral data sets remains a major challenge. In this chapter we presented a model for identifying patterns of social behavior and relating them to environmental and genetic factors, as inspired by machine learning methods for uncovering underlying topics in corpora of documents.

The behavioral states our model discovered and their relationships with demographic covariates were consistent with our knowledge of rhesus macaque behavior generally. Macaques on average spent about 50% of their time in S1–S3, which involved little social interaction beyond receiving non-contact aggression and other agonistic encounters in S1. This pattern may reflect conflict over food or space given that state's high levels of feeding and traveling. The other 50% of the observed time was evenly distributed across the remaining states, S4–S10, which consisted of idiosyncratic combinations of agonistic and affiliative behaviors, reflecting the long-known importance of sociality in rhesus macaque life (Carpenter, 1942; Rawlins & Kessler, 1986). Among both males and females, higher dominance rank was associated with a shift away from the frequent S1 and S3, which involved mostly self-directed behavior amidst some agonistic interactions, and towards S4–S10 which had higher rates of many different social interactions overall as well as high rates of affiliative interactions. This is consistent with the many previous findings that high ranking macaques, and other primates as well, receive higher rates of affiliative behaviors from others in their social group, while low ranking individuals are more likely to experience social isolation (Carpenter, 1942; Brent et al., 2013; Brent, 2015; Seyfarth, 1980, 1977; Bernstein & Sharpe, 1966). The effect of sex on predicted phenotypes was similar to the effect of dominance rank, with males, like low-ranking animals, having higher probabilities of assignment to S1 and S3 at the expense of time spent in S4–S9 (but not S10, notably). This may reflect the inheritance of female rank in rhesus societies, with females remaining in their natal groups and acquiring a rank just below their mother. By contrast, males disperse into new social groups at adulthood and must earn their ranks there. This results in dominance hierarchies generally being more

stable, and therefore social relationships more peaceful, among females than males (Colvin, 1986; Melnick, Pearl, & Richard, 1984).

Our model improves upon the commonly used factor analysis and PCA methods in that it captures information regarding the way individual behaviors co-occur. First, factor analysis and related methods implicitly assumes that data is normally distributed (Roweis & Ghahramani, 1999). This assumption is strongly violated by observations of natural behaviors, which are typically represented as counts. Furthermore, many behaviors of interest are relatively infrequent (see section 2.6 on data processing and likelihood), leading to highly skewed, zero-inflated distributions for individual behaviors. Our model uses instead a flexible discrete distribution to avoid mismatches between the assumed data distribution and the data itself (though any distribution with conjugate priors can be substituted with minimal effort). Second, and more because of the multivariate normal assumption, PCA and factor models can only represent relationships between two behaviors as correlations. The behavioral state model, being a type of mixture model, is more flexible and capable of capturing nonlinear relationships and interaction effects (Gelman, Hwang, & Vehtari, 2013). This means PCA cannot capture nonlinear relationships between pairs of behaviors, or relationships which are modulated by a third behavior. The seemingly quadratic relationship between time spent feeding and traveling in Cayo Santiago macaques, which itself depends on whether aggression is received, is an example of the kind of complex relationship among behaviors that that the behavioral state model is able to represent.

We are not the first to apply topic model-like methods to primate behavioral data. The popular program STRUCTURE, which was in fact developed prior to topic models, represents the genotypes of organisms as distributions over distinct genetic populations (Pritchard, Stephens, & Donnelly, 2000), just as topic models treat documents as distributions over topics and just as our current model describes the phenotype of an animal as a distribution over behavioral states. This model has been applied to distinguish Indian and Chinese rhesus macaque genotypes (Hernandez et al., 2007). The current model is, to the best of

our knowledge, the first attempt to apply topic models to natural animal behavior.

In this chapter we used a pedigree to estimate additive genetic influences on behavioral phenotypes, as per classical heritability analyses. An alternative approach is to include individual genetic variants as predictors in the model as in genome-wide association analyses (GWAS). Our model's regression layer makes GWAS very straightforward to implement, as variant data such as minor allele count can be included directly in the model as a per-animal covariate. However, while this approach would yield much more precise information on how different genetic variants influence phenotypes, it may be difficult to implement for two reasons. First, the traditional GWAS methodology involves fitting a separate model for many common genetic variants (Balding, 2006). While this is tractable when the model being fit is a standard linear regression, fitting a complex hierarchical model thousands of times is computationally infeasible. Second, the sample sizes available in observational data sets of wild or free-ranging animals reach the hundreds or low thousands at best, which are orders of magnitude too small to achieve adequate power to detect the generally small effects of common genetic variants (Chabris et al., 2012; Spencer, Su, Donnelly, Marchini, & Marchini, 2009; Willer et al., 2009). We will return to the latter issue and discuss potential solutions in subsequent chapters.

As neuroscientists and biologists investigate the biological determinants of natural social behavior and attempt to translate laboratory findings into more realistic, unconstrained environments, they face the challenge of quantifying natural social behavior. Because natural social behavior is high-dimensional, highly variable, and yet highly structured, straightforward measurement of its influences is difficult. We have worked to address this issue by developing a model for identifying patterns of social behavior and relating them to environmental and genetic factors. Based on techniques in machine learning for identifying latent structure among documents, this method captures a wider range of relationships among behaviors than is possible using popular methods such as factor analysis or PCA, and explicitly disentangles variability in an individual's social behavior from variability across

the population. We hope that this model will aid researchers in quantifying social behaivior in a way that is rigorous and consistent, while also capturing the richness and complexity of social behavior.

CHAPTER 5 : CONCLUSION

Capturing the full complexity of behavior in social phenotypes

The central aim of the work presented in this thesis has been to bring the representation of social phenotypes in non-human primates closer to the level of the social interaction and the particular patterns of behaviors exhibited in engagement with, or in avoidance of, other agents. This aim was motivated by the presumption that this level of analysis should be more closely connected to neurobiological and physiological causes than the long-run summaries or outcomes of behavior which are commonly studied in both humans and non-human primates. Because the study of non-human primates provides opportunities to observe both the fine structure of social interactions and their long-term consequences, it can help bridge the gap between neurobiological processes, specific instances of social behavior, and long-term social outcomes.

In the previous chapters we have taken a modest step towards quantifying the full complexity of social interactions by developing models which consider not just the occurrence of individual behaviors but also the combinations in which they are deployed. We also found that this step revealed novel information, both qualitatively and quantitatively, regarding the behavioral phenotypes of rhesus macaques. However, there is still much nuance and complexity to rhesus macaque social interactions, known from a long and rich literature from psychology and primatology, that the behavioral strategy models developed here fail to capture. Below I briefly discuss two extensions to the models developed here which bring the representation of behavioral phenotypes closer to the full complexity of social interactions.

Behavioral strategies in continuous time

A major simplification throughout all analyses presented in this thesis is that all behaviors which occur during a focal sample are treated as simultaneous. This makes it difficult to capture the kinds of complex strategies that depend on the particular sequences of events.

Tactical redirection, for example, in which an animal defuses tensions with an antagonistic animal by redirecting their attention, involves receiving an angonistic approach followed by emitting a threat or or vocalization to a third party (Altmann, 1962; Mercier et al., 2017; Wheeler, 2009; Whiten & Byrne, 1988), while the recruitment of coalitions to defend against or perpetrate aggression is indicated by the ordering in which animals approach and attack one another (Altmann, 1962; Silk, 1982; Schülke, Bhagavatula, Vigilant, & Ostner, 2010). These particular sequences of actions may have some signature in terms of the combinations of behavior considered simultaneously that can be captured by the behavioral strategy models developed above, but even if this is true it still would be difficult to determine whether a particular behavioral strategy encompassed something like coalitionary aggression.

Therefore, an important extension to the models presented here would be to consider the occurrence of behavior in real time. The problem of modeling the dependencies between different kinds of data in real time has attracted interest in the Machine Learning literature, which has developed a number of useful methods that may be applicable to NHP social behavior. One such tool is the Hawkes Process, which describes a continuous self-exciting point process in which the occurrence of an event drives the rates of other events in the immediate future (Hawkes, 1971). The Hawkes process has, for example, been used to model social interactions by email between humans where an email from one person excited a response from another (Blundell, Beck, & Heller, 2012; Guo, Blundell, Wallach, & Heller, 2015). Such formulations could be adapted to the domain of interactions between animals by construction a model where the behavior of one animal directed towards another excites or suppresses the rates at which the receiver exhibits different behaviors. A behavioral strategy under this model would then correspond to a particular pattern of excitation between individual behaviors. One challenge in applying this framework would be representing and estimating the differences between individual animals.

Another advantage of a continuous time model is that it would be more generally applicable. The behavioral strategy model developed here are tailored to the particular structure of the

behavioral data collected from Cayo Santiago. The assumption that each focal observation is associated with a single strategy is tenable and supported by the data when each sample is relatively short, but this is not likely to be the case with the hour-long focal samples used in, e.g. Koski (2011). A continuous-time extension of the model would be equally valid for long, short, or continuous observation periods.

Social relationships among rhesus macaques

Extensive research on many species of non-human primates has established that individuals form unique relationships with group members and are aware of and sensitive to the relationships among third parties. Animals form lasting social bonds with others, associated with consistent expression of affiliative or agonistic social interactions over time (Cheney & Seyfarth, 1986; Silk, 2007a; Silk, 2007b). The relationships between animals are also influenced by indirect relationships with third parites; for example an animal will respond differently to the vocalizations of another individual if that individual is a kin or a close associate of an opponent (Cheney, Seyfarth, & Smuts, 1986; Cheney & Seyfarth, 2018; Wittig, Crockford, Seyfarth, & Cheney, 2007; Wittig, Crockford, Langergraber, & Zuberbühler, 2014). These bonds between individuals are influenced, but not determined, by kinship relations and relative dominance rank (Seyfarth et al., 2014; Silk, 1982). In the behavioral state models presented here, we treat relationships between animals as categorical and determined on the basis of kinship or relative dominance rank (the model described in Chapter 3 uses only relative dominance rank, but the outcomes are qualitatively and quantitatively similar whether distinctions are made on kin relationships or dominance rank). Accordingly, the behavioral strategies we estimate cannot take into account the true nature of the relationships between two animals; for instance a strategy of coalitionary aggression or defense among friends against long-term enemies. A natural extension to the behavioral strategy model is therefore to endogenously identify the nature of relationships between animals within the context of the model. For example, a level of categorical latent variables could be used to represent whether each pair of animals had a strong, weak, or antagonistic

relationship, and behavioral strategies then defined on the basis of probabilities of behaviors with animals of each kind of relationship.

Behavioral genetics in free-ranging animal populations

Regardless of the method used to quantify social phenotypes, efforts to identify relationships between genotypic variability and behavior face serious challenges. This topic is of general relevance because genomic association is one of, if not the most used methods for drawing links between neurobiological and physiological processes and behavioral phenotypes measured outside of the laboratory.

As we have noted before, both genome-wide and candidate gene studies of complex phenotypes with small sample sizes have low power to detect genetic effects and low replication rates (Border et al., 2019; Chabris et al., 2012; Ho et al., 2010; Ioannidis et al., 2011; Siontis et al., 2010). In general, the ability to accurately and replicably detect the effects of single common genetic variants will be limited by their effect sizes. The largest reliable reported effect sizes in genome-wide association studies (GWAS) come from studies of morphological phenotypes with extremely high genetic contributions, such as height and body mass index, where the most strongly associated SNVs explained on the order of 0.4% of variation (Speliotes et al., 2010). Smaller single-allele effect sizes have been reported in recent large-scale studies of educational achievement, which is a complex phenotype with a smaller and indirect genetic contribution (Branigan, McCallum, & Freese, 2013). The most predictive SNVs in these studies explain on the order of 0.02% of the total variance (Rietveld et al., 2013; Okbay et al., 2016). Social phenotypes in NHPs are likely to fall somewhere between these two extremes of being directly versus indirectly linked to genetics, and so 0.02% and 0.4% may be reasonable upper and lower bounds for plausible single SNV effect sizes on that phenotype. Such effect sizes imply required sample sizes on the order of thousands in order to reach a power to detect effects above 50%. In human research it is only in the last decade that sample sizes in GWAS have reached a threshold where effects on complex phenotypes can be reliably and replicably detected. It is unlikely that studies of

free-ranging animal populations will achieve these size in the near future.

Despite these difficulties in studying individual genetic variants, natural behavior is predictable on the basis of genetics in both humans and animals. Our results in Chapters 2 and 4 suggest that social behaviors on Cayo Santiago, quantified in terms of combinations of behavior or as rates of individual behaviors, have a modest additive genetic component as measured in terms of heritability. This finding is consistent with previous research on the same macaque population (Brent et al., 2013; Brent et al., 2014), as well as the ubiquitous finding that human phenotypes in general show heritabilities on the order of 0.3 to 0.6 (Polderman et al., 2015; Turkheimer, 2000). This finding is also consistent with the theory that genetic influences on the social behaviors studied here have a massively polygenic genetic architecture, and individual variants have very small effects (Fisher, 1918; Anney et al., 2012; Benjamin, Cesarini, Chabris, et al., 2012; Boyle et al., 2017; Chabris et al., 2013; Davies et al., 2011; Yang et al., 2015; Yang et al., 2010).

Considering the distributed nature of genomic effects, it may be fruitful to broaden the scope of the common variants examined to include not just the variants in a specific gene such as OXTR, but also the broader gene networks that may impact neurobiological function. Recent research suggests that aggregating information across large numbers of common variants may permit the identification of genetic contributions from specific genomic regions, even in sample sizes that are small relative to those used in traditional GWAS research (Benjamin, Cesarini, van der Loos, et al., 2012; Yang, Manolio, et al., 2011). Interestingly, this method is analogous to the method long used in quantitative genetics to estimate heritability on the basis of family relationships, only with genotype data used to estimate genetic similarity rather than kinship; hence this method is sometimes referred to as "realized relatedness" (Powell, Visscher, & Goddard, 2010; Speed et al., 2012). These methods have been used to partition the genomic influence in a trait into the contributions of distinct subsets of genomic regions. Yang, Lee, et al. (2011) found that in several morphological phenotypes, each chromosome explained an amount of trait variability proportional to the

length of the chromosome, while Gusev et al. (2014) found that a large amount of the genomic contribution to several complex disease phenotypes was localized to non-coding regulatory regions. Similar methods could be applied to primate behavior, for example to estimate the contribution of all known neurotransmitter genes or all genes involved in neuron development. However, little is currently known regarding the sample sizes required to reliably estimate the contributions of a gene set or network.

Other avenues for future research are the study of rare, rather than common genetic variants, or gene expression rather than genotype. Rare variants with very low minor allele frequency and *de novo* mutations are likely to have larger effects on complex phenotypes than common variants (Gratten et al., 2014; Neale et al., 2012), and because common variants are imperfectly correlated with rare variants and not at all with *de novo* variants (Eberle et al., 2006; Speed et al., 2012), those sources of genetic variability are not well captured by common genetic variants that are most often studied. Recent studies have also found large effects of social relationships on gene expression (Kohn et al., 2016; Snyder-Mackler et al., 2016), albeit in captive rather than a free-ranging populations.

The novel methods developed in Chapters 3 and 4 of this thesis can make more effective use of observational behavioral data than conventional methods of quantifying social phenotypes from NHP behavior. We found that modeling phenotypes in terms of behavioral strategies increased the information gained about animal phenotypes by around 25%. Providing more information about how animals differ from each other should translate into more efficient estimation and identification of neurobiological correlates of behavior, because it effectively increases the effect size of any covariate that predicts difference between animals, unless the novel information uncovered is completely unrelated to biological factors. However, while the magnitude of the increase in information is not insubstantial, it is also not large enough to alter the reality that the bottleneck for genetic association studies is sample size. It may be more relevant for biological covariates with larger average effect sizes, such as rare variants or gene expression.

Machine learning methods across the behavioral hierarchy

While here we have focused on the modeling and representation of behavioral phenotypes in the context of the primate personality literature, in recent years neuroethologists have taken an increasing interest in fine structure of natural behavior in laboratory animal models. Several lines of research are examining the natural behavior of model organisms such as mice and *Drosophila*, and as in the present work they are increasingly turning to machine-learning methods to aid in parsing that behavior (Branson, Robie, Bender, Perona, & Dickinson, 2009; Geissmann et al., 2017; Reiser, 2009). One popular approach that has been used primarily in *Drosophila* is to use supervised computer-vision algorithms to automatically label behaviors of animals according to a human-defined ethogram (Branson et al., 2009; Kabra, Robie, Rivera-Alba, Branson, & Branson, 2013), while another is to use unsupervised machine-learning models to both label behavior and discover the underlying ethogram (Vogelstein et al., 2014; Wiltschko et al., 2015; Sharma, Johnson, Engert, & Linderman, 2018). The supervised method in Branson et al. (2009), for example, labels individual frames of video data based on a set of predefined physical descriptors of the state of the animal (e.g. posture, velocity). The unsupervised method developed in Vogelstein et al. (2014) learns to categorize and label the short-term reaction of *Drosophila* larvae to stimulation, while the method in Wiltschko et al. (2015) uses a Hidden Markov Model to break down a continuous series of positional data in mice into a sequence of discrete states with subsecond duration, and Sharma et al. (2018) uses a combination of discrete states and continuous trajectories to classify the movement patterns of larval zebrafish.

Our work here differs from the machine learning methods described above in that we are not identifying and labeling individual behaviors. The outputs of the machine-learning algorithms described above are analogous to the focal samples used here as raw data. Similarly, our analog to the machine learning algorithms themselves is the dedicated research staff working on Cayo Santiago who have trained extensively in recognizing different behaviors and labeled their occurrences during in-person observation. Our innovation is in examining how these

elementary behaviors are assembled into temporally-extended patterns reflecting the context and longer-term goals of the animals, and how statistical methods can measure such patterns. To the best of our knowledge, this sort of analysis of in terms of combinations of co-occurring behavior is novel both to animal personality research and to neuroethology. Like animal personality research, neuroethological research using machine learning methods has focused on rates of individual behaviors. Robie et al. (2017), for example, which measured differences in behavior between *Drosophila* genetic lines, represented the behavioral phenotype of each line as a vector of rates of individual behaviors (e.g., walking, chasing, grooming). Similarly, Vogelstein et al. (2014), working with *Drosophila* larvae and examining differences in behavior resulting from the optogenetic activation of different lines of neurons, represented phenotypes as probabilities of each of the reactions identified by their unsupervised learning algorithm.

While it is possible that in *Drosophila*, rates of individual behaviors contain all relevant information about behavioral phenotypes, we doubt this is the case in NHP such as rhesus macaques. The rich documentation of the sophistication of NHP behavior suggests otherwise (Altmann, 1962), as do the results presented here. That the patterns of co-occurrence of individual behaviors may matter is raised in Wiltschko et al. (2015), who note that complex patterns of behaviors are built from simple modules; their work focused on the detection and identification of those simple modules. The present work is complimentary to these methods, in that it attempts to assemble those modules into the overarching complex patterns. Though the particular models developed here are tailored to the focal sampling data widely available for non-human primates, they could also be integrated directly into machine learning algorithms developed in neuroethology.

Bibliography

Agresti, A. (2002). *Categorical data analysis* (2nd). Wiley.

Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679. doi:10.1080/01621459.1993.10476321

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3-4), 227–266. doi:10.1163/156853974X00534

Altmann, S. A. (1962). A field study of the sociobiology of rhesus monkeys, Macaca mulatta. *Annals of the New York Academy of Sciences*, *102*(2), 338–435. doi:10.1111/j.1749-6632.1962.tb13650.x

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th). Arlington, VA: American Psychiatric Association.

Anestis, S. F. (2005). Behavioral style, dominance rank, and urinary cortisol in young chimpanzees (Pan troglodytes). *Behaviour*, *142*(9-10), 1245–1268. doi:10.1163/156853905774539418

Anney, R., Klei, L., Pinto, D., Almeida, J., Bacchelli, E., Baird, G., ... Devlin, B. (2012). Individual common variants exert weak effects on the risk for autism spectrum disorders. *Human Molecular Genetics*, *21*(21), 4781–92. doi:10.1093/hmg/dds301

Apicella, C. L., Cesarini, D., Johannesson, M., Dawes, C. T., Lichtenstein, P., Wallace, B., ... Westberg, L. (2010). No association between oxytocin receptor (OXTR) gene polymorphisms and experimentally elicited social preferences. *PLOS ONE*, *5*(6), e11153. doi:10.1371/journal.pone.0011153

Araya-Ajoy, Y. G., Mathot, K. J., & Dingemanse, N. J. (2015). An approach to estimate short-term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution*, *6*(12), 1462–1473. doi:10.1111/2041-210X.12430

Auyeung, B., Lombardo, M. V., Heinrichs, M., Chakrabarti, B., Sule, A., Deakin, J. B., ... Baron-Cohen, S. (2015). Oxytocin increases eye contact during a real-time, naturalistic social interaction in males with and without autism. *Translational Psychiatry*, *5*(2), e507. doi:10.1038/tp.2014.146

Bakermans-Kranenburg, M. J. & van IJzendoorn, M. H. (2014). A sociability gene? Meta-analysis of oxytocin receptor genotype effects in humans. *Psychiatric Genetics*, *24*(2), 45–51. doi:10.1097/YPG.0b013e3283643684

Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*, *7*(10), 781–91. doi:10.1038/nrg1916

Barak, B. & Feng, G. (2016). Neurobiology of social behavior abnormalities in autism and Williams syndrome. *Nature Neuroscience*, *19*(5), 647–655. doi:10.1038/nn.4276

Bartz, J. A., Zaki, J., Bolger, N., & Ochsner, K. N. (2011). Social effects of oxytocin in humans: Context and person matter. *Trends in Cognitive Sciences*, *15*(7), 301–309. doi:10.1016/j.tics.2011.05.002

Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Gunason, V., ... Lichtenstein, P. (2012). The promises and pitfalls of genoeconomics. *Annual Review of Economics*, *4*, 627–662. doi:10.1146/annurev-economics-080511-110939

Benjamin, D. J., Cesarini, D., van der Loos, M. J. H. M., Dawes, C. T., Koellinger, P. D., Magnusson, P. K. E., ... Visscher, P. M. (2012). The genetic architecture of economic

and political preferences. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(21), 8026–31. doi:10.1073/pnas.1120666109

Bernstein, I. S. & Sharpe, L. G. (1966). Social roles in a rhesus monkey group. *Behaviour*, *26*(1/2), 91–104. doi:10.1163/156853966X00038

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. doi:10.1137/141000671

Bishop, C. (2006). *Pattern recognition and machine learning*. Information Science and Statistics. New York: Springer-Verlag.

Bitetti, M. S. D. (2000). The distribution of grooming among female primates: Testing hypotheses with the shannon-wiener diversity index. *Behaviour*, *137*(11), 1517–1540. doi:10.1163/156853900502709

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77. doi:10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022. Retrieved from http://jmlr.csail.mit.edu/papers/v3/blei03a.html

Blomquist, G. E. (2009). Fitness-related patterns of genetic variation in rhesus macaques. *Genetica*, *135*(2), 209–219. doi:10.1007/s10709-008-9270-x

Blomquist, G. E. & Brent, L. J. N. (2014). Applying quantitative genetic methods to primate social behavior. *International Journal of Primatology*, *35*(1), 108–128. doi:10.1007/s10764-013-9709-5

Blundell, C., Beck, J., & Heller, K. A. (2012). Modelling reciprocating relationships with Hawkes processes. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 25* (pp. 2600–2608). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/4834-modelling-reciprocating-relationships-with-hawkes-processes

Border, R., Johnson, E. C., Evans, L. M., Smolen, A., Berley, N., Sullivan, P. F., & Keller, M. C. (2019). No support for historical candidate gene or candidate gene-by-interaction hypotheses for major depression across multiple large samples. *American Journal of Psychiatry*, *176*(5), 376–387. doi:10.1176/appi.ajp.2018.18070881

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: From polygenic to omnigenic. *Cell*, *169*(7), 1177–1186. doi:10.1016/j.cell.2017.05.038

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social Forces*, *92*(1), 109–140. doi:10.1093/sf/sot076

Branson, K., Robie, A. A., Bender, J., Perona, P., & Dickinson, M. H. (2009). High-throughput ethomics in large groups of *Drosophila*. *Nature Methods*, *6*(6), 451–457. doi:10.1038/nmeth.1328

Brent, L. J. N., Ruiz-Lambides, A., & Platt, M. L. (2017). Family network size and survival across the lifespan of female macaques. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1854), 20170515. doi:10.1098/rspb.2017.0515

Brent, L. J. N. (2010). *Investigating the causes and consequences of sociality in adult female rhesus macaques using a social network approach.* (Doctoral dissertation, University of Roehampton).

Brent, L. J. N. (2015). Friends of friends: Are indirect connections in social networks important to animal behaviour? *Animal Behaviour*. doi:10.1016/j.anbehav.2015.01.020

Brent, L. J. N., Heilbronner, S. R., Horvath, J. E., Gonzalez-Martinez, J., Ruiz-Lambides, A., Robinson, A. G., ... Platt, M. L. (2013). Genetic origins of social networks in rhesus macaques. *Scientific Reports*, *3*, 1042. doi:10.1038/srep01042

Brent, L. J. N., Semple, S., Maclarnon, A., Ruiz-Lambides, A., Gonzalez-Martinez, J., & Platt, M. L. (2014). Personality traits in rhesus macaques (Macaca mulatta) are heritable but do not predict reproductive output. *International Journal of Primatology*, *35*(1), 188–209. doi:10.1007/s10764-013-9724-6

Browning, B. L. & Browning, S. R. (2016). Genotype imputation with millions of reference samples. *The American Journal of Human Genetics*, *98*(1), 116–126. doi:10.1016/j.ajhg.2015.11.020

Budaev, S. V. (2010). Using principal components and factor analysis in animal behaviour research: Caveats and guidelines. *Ethology*, *116*(5), 472–480. doi:10.1111/j.1439-0310.2010.01758.x

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*(1), 23–35. doi:10.1007/s00265-010-1029-6

Burns, T. & Patrick, D. (2007). Social functioning as an outcome measure in schizophrenia studies. *Acta Psychiatrica Scandinavica*, *116*(6), 403–418. doi:10.1111/j.1600-0447.2007.01108.x

Capitanio, J. P. (1999). Personality dimensions in adult male rhesus macaques: Prediction of behaviors across time and situation. *American Journal of Primatology*, *47*(4), 299–320. doi:10.1002/(SICI)1098-2345(1999)47:4⟨299::AID-AJP3⟩3.0.CO;2-P

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. doi:10.18637/jss.v076.i01

Carpenter, C. R. (1942). Characteristics of social behavior in non-human primates. *Transactions of the New York Academy of Sciences*, *4*(8 Series II), 248–258. doi:10.1111/j.2164-0947.1942.tb00856.x

Chabris, C. F., Lee, J. J., Benjamin, D. J., Beauchamp, J. P., Glaeser, E. L., Borst, G., ... Laibson, D. I. (2013). Why it is hard to find genes associated with social science traits: Theoretical and empirical considerations. *American Journal of Public Health*, *103*(S1), S152–S166. doi:10.2105/AJPH.2013.301327

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., ... Laibson, D. (2012). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science*, *23*(11), 1314–23. doi:10.1177/0956797611435528

Chang, S. W. C., Fagan, N. A., Toda, K., Utevsky, A. V., Pearson, J. M., & Platt, M. L. (2015). Neural mechanisms of social decision-making in the primate amygdala. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(52), 16012–16017. doi:10.1073/pnas.1514761112

Charpentier, M. J. E., Prugnolle, F., Gimenez, O., & Widdig, A. (2008). Genetic heterozygosity and sociality in a primate species. *Behavior Genetics*, *38*(2), 151–158. doi:10.1007/s10519-008-9191-6

Chen, J., Zhu, J., Wang, Z., Zheng, X., & Zhang, B. (2013). Scalable inference for logistic-normal topic models. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 26* (pp. 2445–2453). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/4981-scalable-inference-for-logistic-normal-topic-models

Cheney, D., Seyfarth, R., & Smuts, B. (1986). Social relationships and social cognition in nonhuman primates. *Science*, *234*(4782), 1361–1366. doi:10.1126/science.3538419

Cheney, D. L. & Seyfarth, R. M. (1986). The recognition of social alliances by vervet monkeys. *Animal Behaviour*, *34*(6), 1722–1731. doi:10.1016/S0003-3472(86)80259-7

Cheney, D. L. & Seyfarth, R. M. (2018). Flexible usage and social function in primate vocalizations. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9), 1974–1979. doi:10.1073/pnas.1717572115

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., ... Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, *6*(2), 80–92. doi:10.4161/fly.19695

Colvin, J. D. (1986). Proximate causes of male emigration at puberty in rhesus monkeys. In R. G. Rawlins & M. J. Kessler (Eds.), *The Cayo Santiago macaques: History, behavior, and biology*. State University of New York Press.

Costa, B., Pini, S., Gabelloni, P., Abelli, M., Lari, L., Cardini, A., ... Martini, C. (2009). Oxytocin receptor polymorphisms and adult attachment style in patients with depression. *Psychoneuroendocrinology*, *34*(10), 1506–1514. doi:10.1016/j.psyneuen.2009.05.006

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied Statistics*, *21*(2), 113. doi:10.2307/2346482

Cox, D. R. & Wermuth, N. (1994). A note on the quadratic exponential binary distribution. *Biometrika*, *81*(2), 403. doi:10.2307/2336971

Crawley, J. N. (2007). *What's wrong with my mouse? Behavioral phenotyping of transgenic and knockout mice* (2nd). Hoboken, NJ, US: John Wiley & Sons Inc.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., ... Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*(10), 996–1005. doi:10.1038/mp.2011.85

Davies, S. W., Scarpino, S. V., Pongwarin, T., Scott, J., & Matz, M. V. (2015). Estimating trait heritability in highly fecund species. *G3: Genes, Genomes, Genetics*, *5*(12), 2639–45. doi:10.1534/g3.115.020701

Dempster, E. L., Burcescu, I., Wigg, K., Kiss, E., Baji, I., Gadoros, J., ... Barr, C. L. (2007). Evidence of an association between the vasopressin V1b receptor gene (AVPR1b) and childhood-onset mood disorders. *Archives of General Psychiatry*, *64*(10), 1189. doi:10.1001/archpsyc.64.10.1189

Dingemanse, N. J. & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. doi:10.1111/1365-2656.12013

Eager, C. D. (2017). Standardize: Tools for standardizing variables for regression in R. Retrieved from https://CRAN.R-project.org/package=standardize

Eberle, M. A., Rieder, M. J., Kruglyak, L., & Nickerson, D. A. (2006). Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome. *PLOS Genetics*, *2*(9), e142. doi:10.1371/journal.pgen.0020142

Ebstein, R. P., Knafo, A., Mankuta, D., Chew, S. H., & Lai, P. S. (2012). The contributions of oxytocin and vasopressin pathway genes to human behavior. *Hormones and Behavior*, *61*(3), 359–379. doi:10.1016/j.yhbeh.2011.12.014

Efron, B. (1978). Regression and ANOVA with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, *73*(361), 113–121. doi:10.2307/2286531

Fedigan, L. M. (1983). Dominance and reproductive success in primates. *American Journal of Physical Anthropology*, *26*(S1), 91–129. doi:10.1002/ajpa.1330260506

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. doi:10.1017/S0080456800012163

Fowler, J. H., Baker, L. A., & Dawes, C. T. (2008). Genetic variation in political participation. *American Political Science Review*, *102*(2), 233–248. doi:10.1017/S0003055408080209

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. doi:10.1002/ajp.20833

Freeman, S. M., Inoue, K., Smith, A. L., Goodman, M. M., & Young, L. J. (2014). The neuroanatomical distribution of oxytocin receptor binding and mRNA in the male rhesus macaque (Macaca mulatta). *Psychoneuroendocrinology*, *45*, 128–141. doi:10.1016/j.psyneuen.2014.03.023

Geissmann, Q., Rodriguez, L. G., Beckwith, E. J., French, A. S., Jamasb, A. R., & Gilestro, G. F. (2017). Ethoscopes: An open platform for high-throughput ethomics. *PLOS Biology*, *15*(10), e2003026. doi:10.1371/journal.pbio.2003026

Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534. doi:10.1214/06-BA117A

Gelman, A. & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, *9*(6), 641–651. doi:10.1177/1745691614551642

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189–211. doi:10.1080/19345747.2011.618213

Gelman, A., Hwang, J., & Vehtari, A. (2013). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016. doi:10.1007/s11222-013-9416-2

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. doi:10.1214/08-AOAS191

Gibbs, R. A., Rogers, J., Katze, M. G., Bumgarner, R., Weinstock, G. M., Mardis, E. R., ... Zwieg, A. S. (2007). Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, *316*(5822), 222–34. doi:10.1126/science.1139247

Glusman, G., Caballero, J., Mauldin, D. E., Hood, L., & Roach, J. C. (2011). Kaviar: An accessible system for testing SNV novelty. *Bioinformatics*, *27*(22), 3216–3217. doi:10.1093/bioinformatics/btr540

Gosling, S. D. (2001). From mice to men: What can we learn about personality from animal research? *Psychological Bulletin*, *127*(1), 45–86. doi:10.1037/0033-2909.127.1.45

Gratten, J., Wray, N. R., Keller, M. C., & Visscher, P. M. (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. *Nature Neuroscience*, *17*(6), 782–790. doi:10.1038/nn.3708

Guo, F., Blundell, C., Wallach, H., & Heller, K. (2015). The Bayesian Echo Chamber: Modeling social influence via linguistic accommodation. In *Artificial Intelligence and Statistics* (pp. 315–323). Retrieved from http://proceedings.mlr.press/v38/guo15.html

Gusev, A., Lee, S. H., Trynka, G., Finucane, H., Vilhjálmsson, B. J., Xu, H., ... Sullivan, P. F. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics*, *95*(5), 535–552. doi:10.1016/j.ajhg.2014.10.004

Haberman, S. J. (1982). Analysis of dispersion of multinomial responses. *Journal of the American Statistical Association*, *77*(379), 568. doi:10.2307/2287713

Harden, K. P. (2014). Genetic influences on adolescent sexual behavior: Why genes matter for environmentally-oriented researchers. *Psychological Bulletin*, *140*(2), 434–465. doi:10.1037/a0033564

Hart, A. B., de Wit, H., & Palmer, A. A. (2013). Candidate gene studies of a promising intermediate phenotype: Failure to replicate. *Neuropsychopharmacology*, *38*(5), 802–816. doi:10.1038/npp.2012.245

Hashikawa, K., Hashikawa, Y., Falkner, A., & Lin, D. (2016). The neural circuits of mating and fighting in male mice. *Current Opinion in Neurobiology*. Neurobiology of Sex, *38*, 27–37. doi:10.1016/j.conb.2016.01.006

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Series in Statistics. Springer.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*(1), 83–90. doi:10.1093/biomet/58.1.83

Hecht, M., Bromberg, Y., & Rost, B. (2015). Better prediction of functional effects for sequence variants. *BMC Genomics*, *16*(8), S1. doi:10.1186/1471-2164-16-S8-S1

Henderson, C. R., Kempthorne, O., Searle, S. R., & von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, *15*(2), 192. doi:10.2307/2527669

Hernandez, R. D., Hubisz, M. J., Wheeler, D. A., Smith, D. G., Ferguson, B., Rogers, J., ... Bustamante, C. D. (2007). Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science*, *316*(5822), 240–3. doi:10.1126/science.1140462

Higham, J. P., Barr, C. S., Hoffman, C. L., Mandalaywala, T. M., Parker, K. J., & Maestripieri, D. (2011). Mu-opioid receptor (OPRM1) variation, oxytocin levels and maternal attachment in free-ranging rhesus macaques Macaca mulatta. *Behavioral Neuroscience*, *125*(2), 131–136. doi:10.1037/a0022695

Hill, D. A. (1999). Effects of provisioning on the social behaviour of Japanese and rhesus macaques: Implications for socioecology. *Primates*, *40*(1), 187–198. doi:10.1007/BF02557710

Ho, M. K., Goldman, D., Heinz, A., Kaprio, J., Kreek, M. J., Li, M. D., ... Tyndale, R. F. (2010). Breaking barriers in the genomics and pharmacogenetics of drug addiction. *Clinical Pharmacology & Therapeutics*, *88*(6), 779–791. doi:10.1038/clpt.2010.175

Hopkins, W. D., Donaldson, Z. R., & Young, L. J. (2012). A polymorphic indel containing the RS3 microsatellite in the 5' flanking region of the vasopressin V1a receptor gene is associated with chimpanzee (Pan troglodytes) personality. *Genes, Brain and Behavior*, *11*(5), 552–558. doi:10.1111/j.1601-183X.2012.00799.x

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science*, *241*(4865), 540–545. doi:10.1126/science.3399889

Huang, A. & Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, *8*(2), 439–452. doi:10.1214/13-BA815

Insel, T. R. (2010). The challenge of translation in social neuroscience: A review of oxytocin, vasopressin, and affiliative behavior. *Neuron*, *65*(6), 768–79. doi:10.1016/j.neuron.2010.03.005

Ioannidis, J., Tarone, R., & McLaughlin, J. (2011). The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology*, *22*(4), 450–456. doi:10.1097/EDE.0b013e31821b506e

Iqbal, S. N., Yin, L., Drucker, C. B., Kuang, Q., Gariépy, J.-F., Platt, M. L., & Pearson, J. M. (2019). Latent goal models for dynamic strategic interaction. *PLOS Computational Biology*, *15*(3), e1006895. doi:10.1371/journal.pcbi.1006895

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Reibold, M., Bachner-Melman, R., ... Ebstein, R. P. (2008). Molecular genetic studies of the arginine vasopressin 1a receptor (AVPR1a) and the oxytocin receptor (OXTR) in human behaviour: From autism to altruism with some notes in between. *Progress in Brain Research*, *170*, 435–49. doi:10.1016/S0079-6123(08)00434-2

Israel, S., Lerer, E., Shalev, I., Uzefovsky, F., Riebold, M., Laiba, E., ... Ebstein, R. P. (2009). The oxytocin receptor (OXTR) contributes to prosocial fund allocations in the Dictator Game and the Social Value Orientations task. *PLOS ONE*, *4*(5), e5535. doi:10.1371/journal.pone.0005535

Jang, K. L., Livesley, W. J., & Vernon, P. A. (1996). Heritability of the big five personality dimensions and their facets: A twin study. *Journal of Personality*, *64*(3), 577–591.

Johansson, A., Bergman, H., Corander, J., Waldman, I. D., Karrani, N., Salo, B., ... Westberg, L. (2012). Alcohol and aggressive behavior in men-moderating effects of oxytocin receptor gene (OXTR) polymorphisms. *Genes, Brain and Behavior*, *11*(2), 214–221. doi:10.1111/j.1601-183X.2011.00744.x

Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S., & Branson, K. (2013). JAABA: Interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, *10*(1), 64–67. doi:10.1038/nmeth.2281

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi:10.1080/01621459.1995.10476572

Keck, M. E., Kern, N., Erhardt, A., Unschuld, P. G., Ising, M., Salyakina, D., ... Binder, E. B. (2008). Combined effects of exonic polymorphisms in CRHR1 and AVPR1B

genes in a case/control study for panic disorder. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *147B*(7), 1196–1204. doi:10.1002/ajmg.b.30750

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, a. D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. doi:10.1101/gr.229102

King, J. E. & Figueredo, A. J. (1997). The five-factor model plus dominance in chimpanzee personality. *Journal of Research in Personality*, *31*(2), 257–271. doi:10.1006/jrpe.1997.2179

Knafo, A., Israel, S., Darvasi, A., Bachner-Melman, R., Uzefovsky, F., Cohen, L., ... Ebstein, R. P. (2008). Individual differences in allocation of funds in the dictator game associated with length of the arginine vasopressin 1a receptor RS3 promoter region and correlation between RS3 length and hippocampal mRNA. *Genes, Brain and Behavior*, *7*(3), 266–275. doi:10.1111/j.1601-183X.2007.00341.x

Kohn, J. N., Snyder-Mackler, N., Barreiro, L. B., Johnson, Z. P., Tung, J., & Wilson, M. E. (2016). Dominance rank causally affects personality and glucocorticoid regulation in female rhesus macaques. *Psychoneuroendocrinology*, *74*, 179–188. doi:10.1016/j.psyneuen.2016.09.005

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U., & Fehr, E. (2005). Oxytocin increases trust in humans. *Nature*, *435*(7042), 673–676. doi:10.1038/nature03701

Koski, S. E. (2011). Social personality traits in chimpanzees: Temporal stability and structure of behaviourally assessed personality traits in three captive populations. *Behavioral Ecology and Sociobiology*, *65*(11), 2161–2174. doi:10.1007/s00265-011-1224-0

Kozorovitskiy, Y., Hughes, M., Lee, K., & Gould, E. (2006). Fatherhood affects dendritic spines and vasopressin V1a receptors in the primate prefrontal cortex. *Nature Neuroscience*, *9*(9), 1094–1095. doi:10.1038/nn1753

Kruk, M. R. (1991). Ethology and pharmacology of hypothalamic aggression in the rat. *Neuroscience & Biobehavioral Reviews*. Special Issue Advances in Ethopharmacology, *15*(4), 527–538. doi:10.1016/S0149-7634(05)80144-7

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the "animal model". *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *359*(1446), 873–90. doi:10.1098/rstb.2003.1437

Lafferty, J. D. & Blei, D. M. (2006). Correlated topic models. In Y. Weiss, P. B. Schölkopf, & J. C. Platt (Eds.), *Advances in Neural Information Processing Systems 18* (pp. 147–154). MIT Press.

Lango Allen, H., Estrada, K., Lettre, G., Berndt, S. I., Weedon, M. N., Rivadeneira, F., ... Hirschhorn, J. N. (2010). Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, *467*(7317), 832–8. doi:10.1038/nature09410

Latzman, R. D., Hopkins, W. D., Keebaugh, A. C., & Young, L. J. (2014). Personality in chimpanzees (Pan troglodytes): Exploring the hierarchical structure and associations with the vasopressin V1A receptor gene. *PLOS ONE*, *9*(4), e95741. doi:10.1371/journal.pone.0095741

Lemoine, N. P., Hoffman, A., Felton, A. J., Baur, L., Chaves, F., Gray, J., ... Smith, M. D. (2016). Underappreciated problems of low replication in ecological field studies. *Ecology*, *97*(10), 2554–2561. doi:10.1002/ecy.1506

Lerer, E., Levi, S., Salomon, S., Darvasi, A., Yirmiya, N., & Ebstein, R. P. (2008). Association between the oxytocin receptor (OXTR) gene and autism: Relationship to Vineland Adaptive Behavior Scales and cognition. *Molecular Psychiatry*, *13*(10), 980–8. doi:10.1038/sj.mp.4002087

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. doi:10.1016/j.jmva.2009.04.008

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–60. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–9. doi:10.1093/bioinformatics/btp352

Li, J., Zhao, Y., Li, R., Broster, L. S., Zhou, C., & Yang, S. (2015). Association of oxytocin receptor gene (OXTR) rs53576 polymorphism with sociality: A meta-analysis. *PLOS ONE*, *10*(6), e0131820. doi:10.1371/journal.pone.0131820

Madlon-Kay, S., Montague, M. J., Brent, L. J. N., Ellis, S., Zhong, B., Snyder-Mackler, N., ... Platt, M. L. (2018). Weak effects of common genetic variation in oxytocin and vasopressin receptor genes on rhesus macaque social behavior. *American Journal of Primatology*, *80*(10). doi:10.1002/ajp.22873

Martin, J. S. & Suarez, S. A. (2017). Personality assessment and model comparison with behavioral data: A statistical framework and empirical demonstration with bonobos (Pan paniscus). *American Journal of Primatology*, *79*(8), e22670. doi:10.1002/ajp.22670

Martin, N. G., Eaves, L. J., Heath, A. C., Jardine, R., Feingold, L. M., & Eysenck, H. J. (1986). Transmission of social attitudes. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(12), 4364–4368. doi:10.1073/pnas.83.12.4364

Maslow, A. H. (1936). The role of dominance in the social and sexual behavior of infra-human primates: IV. The determination of hierarchy in pairs and in groups. *The Pedagogical Seminary and Journal of Genetic Psychology*, *49*, 161–198. doi:10.1080/08856559.1936.10533757

Massen, J. J. & Koski, S. E. (2014). Chimps of a feather sit together: Chimpanzee friendships are based on homophily in personality. *Evolution and Human Behavior*, *35*(1), 1–8. doi:10.1016/j.evolhumbehav.2013.08.008

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. doi:10.1101/gr.107524.110

Melnick, D. J., Pearl, M. C., & Richard, A. F. (1984). Male migration and inbreeding avoidance in wild rhesus monkeys. *American Journal of Primatology*, *7*(3), 229–243. doi:10.1002/ajp.1350070303

Mercier, S., Neumann, C., van de Waal, E., Chollet, E., de Bellefon, J. M., & Zuberbühler, K. (2017). Vervet monkeys greet adult males during high-risk situations. *Animal Behaviour*, *132*, 229–245. doi:https://doi.org/10.1016/j.anbehav.2017.07.021

Meuwissen, T. H., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–29. doi:10.1534/genetics.107.081190

Morgan, K. N., Tromborg, C. T., Syaadah, O., Norma, M., Feldon, J., Berckmans, D., ... Lindburg, D. (2007). Sources of stress in captivity. *Applied Animal Behaviour Science*, *102*(3-4), 262–302. doi:10.1016/j.applanim.2006.05.032

Mori, A. (1977). Intra-troop spacing mechanism of the wild Japanese monkeys of the Koshima troop. *Primates*, *18*(2), 331–357. doi:10.1007/BF02383112

Munk, A. J. L., Hermann, A., El Shazly, J., Grant, P., & Hennig, J. (2016). The idea is good, but...: Failure to replicate associations of oxytocinergic polymorphisms with face-inversion in the N170. *PLOS ONE*, *11*(3), e0151991. doi:10.1371/journal.pone.0151991

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. The MIT Press.

Nakagawa, S. & Schielzeth, H. (2010). Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biological Reviews*, *85*(4), no–no. doi:10.1111/j.1469-185X.2010.00141.x

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., ... Daly, M. J. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, *485*(7397), 242–5. doi:10.1038/nature11011

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

Neumann, C., Agil, M., Widdig, A., & Engelhardt, A. (2013). Personality of wild male crested macaques (Macaca nigra). *PLoS ONE*, *8*(8), e69383. doi:10.1371/journal.pone.0069383

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., ... Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. doi:10.1038/nature17671

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). doi:10.1126/science.aac4716

Park, T. & Min, S. (2014). Partially collapsed Gibbs sampling for linear mixed-effects models. *Communications in Statistics - Simulation and Computation*, *45*(1), 165–180. doi:10.1080/03610918.2013.857687

Parr, L. A., Modi, M., Siebert, E., & Young, L. J. (2013). Intranasal oxytocin selectively attenuates rhesus monkeys' attention to negative facial expressions. *Psychoneuroendocrinology*, *38*(9), 1748–1756. doi:10.1016/j.psyneuen.2013.02.011

Pearson, J. M., Watson, K. K., & Platt, M. L. (2014). Decision making: The neuroethological turn. *Neuron*, *82*(5), 950–965. doi:10.1016/j.neuron.2014.04.037

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. doi:10.1038/ng.3285

Polson, N. G., Scott, J. G., & Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, *108*(504), 1339–1349. doi:10.1080/01621459.2013.829001

Powell, J. E., Visscher, P. M., & Goddard, M. E. (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics*, *11*(11), 800–5. doi:10.1038/nrg2865

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959. Retrieved from http://www.genetics.org/content/155/2/945.long

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163. doi:10.2307/271063

Rawlins, R. G. & Kessler, M. J. (1986). *The Cayo Santiago macaques: History, behavior, and biology*. State University of New York Press.

Reiser, M. (2009). The ethomics era? *Nature Methods*, *6*(6), 413–414. doi:10.1038/nmeth0609-413

Resende, M. F. R., Muñoz, P., Resende, M. D. V., Garrick, D. J., Fernando, R. L., Davis, J. M., ... Kirst, M. (2012). Accuracy of genomic selection methods in a standard data set of loblolly pine (Pinus taeda L.) *Genetics*, *190*(4), 1503–10. doi:10.1534/genetics.111.137026

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., ... Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13790–13794. doi:10.1073/pnas.1404623111

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. G. N. W., ... Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*(6139), 1467–71. doi:10.1126/science.1235488

Robie, A. A., Hirokawa, J., Edwards, A. W., Umayam, L. A., Lee, A., Phillips, M. L., ... Branson, K. (2017). Mapping the neural substrates of behavior. *Cell*, *170*(2), 393–406.e28. doi:10.1016/J.CELL.2017.06.032

Roche, D. G., Careau, V., & Binning, S. A. (2016). Demystifying animal 'personality' (or not): Why individual variation matters to experimental biologists. *The Journal of Experimental Biology*, *219*(24). doi:10.1242/jeb.146712

Rodrigues, S. M., Saslow, L. R., Garcia, N., John, O. P., & Keltner, D. (2009). Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21437–41. doi:10.1073/pnas.0909579106

Rodríguez, A. & Dunson, D. B. (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Analysis*, *6*(1), 145–177. doi:10.1214/11-BA605

Roweis, S. & Ghahramani, Z. (1999). A unifying review of linear Gaussian models. *Neural Computation*, *11*(2), 305–345. doi:10.1162/089976699300016674

Schino, G. (2001). Grooming, competition and social rank among female primates: A meta-analysis. *Animal Behaviour*, *62*(2), 265–271. doi:10.1006/ANBE.2001.1750

Schülke, O., Bhagavatula, J., Vigilant, L., & Ostner, J. (2010). Social bonds enhance reproductive success in male macaques. *Current Biology*, *20*(24), 2207–2210. doi:10.1016/j.cub.2010.10.058

Seyfarth, R. M. (1977). A model of social grooming among adult female monkeys. *Journal of Theoretical Biology*, *65*(4), 671–698. doi:10.1016/0022-5193(77)90015-7

Seyfarth, R. M. (1980). The distribution of grooming and related behaviours among adult female vervet monkeys. *Animal Behaviour*, *28*(3), 798–813. doi:10.1016/S0003-3472(80)80140-0

Seyfarth, R. M., Silk, J. B., & Cheney, D. L. (2014). Social bonds in female baboons: The interaction between personality, kinship and rank. *Animal Behaviour*, *87*, 23–29. doi:10.1016/j.anbehav.2013.10.008

Seyfarth, R. M., Silk, J. B., & Cheney, D. L. (2012). Variation in personality and fitness in wild female baboons. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(42), 16980–5. doi:10.1073/pnas.1210780109

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x

Sharma, A., Johnson, R., Engert, F., & Linderman, S. (2018). Point process latent variable models of larval zebrafish behavior. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 31* (pp. 10919–10930). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/8289-point-process-latent-variable-models-of-larval-zebrafish-behavior

Sherry, S. T., Ward, M.-H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311. doi:10.1093/nar/29.1.308

Silk, J. B. (2007a). Social components of fitness in primate groups. *Science*, *317*(5843), 1347–51. doi:10.1126/science.1140734

Silk, J. B. (2007b). The adaptive value of sociality in mammalian groups. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1480), 539–59. doi:10.1098/rstb.2006.1994

Silk, J. B. (1982). Altruism among female Macaca radiata: Explanations and analysis of patterns of grooming and coalition formation. *Behaviour*, *79*(2-4), 162–188. doi:10.1163/156853982X00238

Silk, J. B. (1999). Male bonnet macaques use information about third-party rank relationships to recruit allies. *Animal Behaviour*, *58*(1), 45–51. doi:10.1006/anbe.1999.1129

Silk, J. B., Alberts, S. C., & Altmann, J. (2003). Social bonds of female baboons enhance infant survival. *Science*, *302*(5648), 1231–4. doi:10.1126/science.1088580

Sinnwell, J. P., Therneau, T. M., & Schaid, D. J. (2014). The kinship2 R package for pedigree data. *Human Heredity*, *78*(2), 91–3. doi:10.1159/000363105

Siontis, K. C. M., Patsopoulos, N. A., & Ioannidis, J. P. A. (2010). Replication of past candidate loci for common diseases and phenotypes in 100 genome-wide association studies. *European Journal of Human Genetics*, *18*(7), 832–837. doi:10.1038/ejhg.2010.26

Smith, A. S., Ågmo, A., Birnie, A. K., & French, J. A. (2010). Manipulation of the oxytocin system alters social behavior and attraction in pair-bonding primates, Callithrix penicillata. *Hormones and Behavior*, *57*(2), 255–262. doi:10.1016/j.yhbeh.2009.12.004

Snowdon, C. T., Pieper, B. A., Boe, C. Y., Cronin, K. A., Kurian, A. V., & Ziegler, T. E. (2010). Variation in oxytocin is related to variation in affiliative behavior in monogamous, pairbonded tamarins. *Hormones and Behavior*, *58*(4), 614–618. doi:10.1016/j.yhbeh.2010.06.014

Snyder-Mackler, N., Sanz, J., Kohn, J. N., Brinkworth, J. F., Morrow, S., Shaver, A. O., ... Tung, J. (2016). Social status alters immune regulation and response to infection in macaques. *Science*, *354*(6315), 1041–1045. doi:10.1126/science.aah3580

Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics*, *91*(6), 1011–1021. doi:10.1016/j.ajhg.2012.10.010

Speliotes, E. K., Willer, C. J., Berndt, S. I., Monda, K. L., Thorleifsson, G., Jackson, A. U., … Loos, R. J. F. (2010). Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nature Genetics*, *42*(11), 937–48. doi:10.1038/ng.686

Spencer, C. C. A., Su, Z., Donnelly, P., Marchini, J., & Marchini, J. (2009). Designing genome-wide association studies: Sample size, power, imputation, and the choice of genotyping chip. *PLOS Genetics*, *5*(5), e1000477. doi:10.1371/journal.pgen.1000477

Staes, N., Koski, S. E., Helsen, P., Fransen, E., Eens, M., & Stevens, J. M. (2015). Chimpanzee sociability is associated with vasopressin (Avpr1a) but not oxytocin receptor gene (OXTR) variation. *Hormones and Behavior*, *75*, 84–90. doi:10.1016/J.YHBEH.2015.08.006

Staes, N., Weiss, A., Helsen, P., Korody, M., Eens, M., & Stevens, J. M. (2016). Bonobo personality traits are heritable and associated with vasopressin receptor gene 1a variation. *Scientific Reports*, *6*(1), 38193. doi:10.1038/srep38193

Steptoe, A., Shankar, A., Demakakos, P., & Wardle, J. (2013). Social isolation, loneliness, and all-cause mortality in older men and women. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5797–5801. doi:10.1073/pnas.1219686110

Stevenson-Hinde, J. & Zunz, M. (1978). Subjective assessment of individual rhesus monkeys. *Primates*, *19*(3), 473–482. doi:10.1007/BF02373309

Theil, H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*. doi:10.2307/2775440

Thierry, B. (1985). Patterns of agonistic interactions in three species of macaque (Macaca mulatta, M fascicularis, M tonkeana). *Aggressive Behavior*, *11*(3), 223–233. doi:10.1002/1098-2337(1985)11:3⟨223::AID-AB2480110305⟩3.0.CO;2-A

Trefilov, A., Berard, J., Krawczak, M., & Schmidtke, J. (2000). Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation. *Behavior Genetics*, *30*(4), 295–301. doi:10.1023/A:1026597300525

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*(5), 160–164. doi:10.1111/1467-8721.00084

Umberson, D. & Karas Montez, J. (2010). Social relationships and health: A flashpoint for health policy. *Journal of Health and Social Behavior*, *51*(1_suppl), S54–S66. doi:10.1177/0022146510383501

van Dyk, D. A. & Park, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, *103*(482), 790–796. doi:10.1198/016214508000000409

Vazquez, A., Weigel, K., Gianola, D., Bates, D., Perez-Cabal, M., Rosa, G., & Chang, Y. (2009). Poisson versus threshold models for genetic analysis of clinical mastitis in US Holsteins. *Journal of Dairy Science*, *92*(10), 5239–5247. doi:10.3168/jds.2009-2085

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. doi:10.1007/s11222-016-9696-4

Vogelstein, J. T., Park, Y., Ohyama, T., Kerr, R. A., Truman, J. W., Priebe, C. E., & Zlatic, M. (2014). Discovery of brainwide neural-behavioral maps via multiscale unsupervised structure learning. *Science*, *344*(6182), 386–392. doi:10.1126/SCIENCE.1250298

Waal, F. B. M. D. & Luttrell, L. M. (1989). Toward a comparative socioecology of the genus Macaca: Different dominance styles in rhesus and stumptail monkeys. *American Journal of Primatology*, *19*(2), 83–109. doi:10.1002/ajp.1350190203

Walum, H., Westberg, L., Henningsson, S., Neiderhiser, J. M., Reiss, D., Igl, W., ... Lichtenstein, P. (2008). Genetic variation in the vasopressin receptor 1a gene (AVPR1A) associates with pair-bonding behavior in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 14153–6. doi:10.1073/pnas.0803081105

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Wheeler, B. C. (2009). Monkeys crying wolf? Tufted capuchin monkeys use anti-predator calls to usurp resources from conspecifics. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1669), 3013–3018. doi:10.1098/rspb.2009.0544

Whiten, A. & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences*, *11*(2), 233–244. doi:10.1017/S0140525X00049682

Widdig, A., Bercovitch, F. B., Streich, W. J., Sauermann, U., Nürnberg, P., & Krawczak, M. (2004). A longitudinal analysis of reproductive skew in male rhesus macaques. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1541), 819–26. doi:10.1098/rspb.2003.2666

Widdig, A., Kessler, M. J., Bercovitch, F. B., Berard, J. D., Duggleby, C., Nürnberg, P., ... Schmidtke, J. (2016). Genetic studies on the Cayo Santiago rhesus macaques: A review of 40 years of research. *American Journal of Primatology*, *78*(1), 44–62. doi:10.1002/ajp.22424

Willer, C. J., Speliotes, E. K., Loos, R. J. F., Li, S., Lindgren, C. M., Heid, I. M., ... Hirschhorn, J. N. (2009). Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nature Genetics*, *41*(1), 25–34. doi:10.1038/ng.287

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., ... Nussey, D. H. (2010). An ecologist's guide to the animal model. *The Journal of Animal Ecology*, *79*(1), 13–26. doi:10.1111/j.1365-2656.2009.01639.x

Wilson, V. A. D., Weiss, A., Humle, T., Morimura, N., Udono, T., Idani, G., ... Inoue-Murayama, M. (2017). Chimpanzee personality and the arginine vasopressin receptor 1A genotype. *Behavior Genetics*, *47*(2), 215–226. doi:10.1007/s10519-016-9822-2

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., ... Datta, S. R. (2015). Mapping sub-second structure in mouse behavior. *Neuron*, *88*(6), 1121–1135. doi:10.1016/J.NEURON.2015.11.031

Wittig, R. M., Crockford, C., Langergraber, K. E., & Zuberbühler, K. (2014). Triadic social interactions operate across time: A field experiment with wild chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1779), 20133155. doi:10.1098/rspb.2013.3155

Wittig, R. M., Crockford, C., Seyfarth, R. M., & Cheney, D. L. (2007). Vocal alliances in Chacma baboons (Papio hamadryas ursinus). *Behavioral Ecology and Sociobiology*, *61*(6), 899–909. doi:10.1007/s00265-006-0319-5

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., ... Visscher, P. M. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*, *47*(10), 1114–20. doi:10.1038/ng.3390

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., ... Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–9. doi:10.1038/ng.608

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. doi:10.1016/j.ajhg.2010.11.011

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., ... Visscher, P. M. (2011). Genome partitioning of genetic variation for complex traits using common SNPs. *Nature Genetics*, *43*(6), 519–525. doi:10.1038/ng.823

Yin, A., Tseng, P. H., Rajangam, S., Lebedev, M. A., & Nicolelis, M. a. L. (2018). Place cell-like activity in the primary sensorimotor and premotor cortex during monkey whole-body navigation. *Scientific Reports*, *8*(1), 1–17. doi:10.1038/s41598-018-27472-4

Young, L. J. & Flanagan-Cato, L. M. (2012). Editorial comment: Oxytocin, vasopressin and social behavior. *Hormones and Behavior*, *61*(3), 227–229. doi:https://doi.org/10.1016/j.yhbeh.2012.02.019

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLOS Genetics*, *9*(2), e1003264. doi:10.1371/journal.pgen.1003264

Zimin, A. V., Cornish, A. S., Maudhoo, M. D., Gibbs, R. M., Zhang, X., Pandey, S., ... Norgren, R. B. (2014). A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biology Direct*, *9*(1), 20. doi:10.1186/1745-6150-9-20