

# Reconciling Rigor and Range: Observations, Experiments, and Quasi-experiments in Field Primatology

Charles H. Janson

Received: 5 June 2011 / Accepted: 17 August 2011 / Published online: 10 November 2011  
© Springer Science+Business Media, LLC 2011

**Abstract** I summarize here my experiences studying biological mechanisms of primate behavior across more than 3 decades of observing behaviors and performing ecological field experiments on a wild primate population. I begin with a general overview of the purpose of controlling variation in predictor variables, then describe how observational and experimental studies control such variation. After a description of the practical problems of using observational data to infer mechanisms or causes of a particular behavior, I describe the success and challenges of conducting ecological field experiments on animals, including some examples of (my own) failed experiments. Finally, I propose a middle ground between observations and experiments, what I term quasi-experiments. Quasi-experiments use systematic observation of predictor variables to measure their effects on the occurrence of a behavior, rather than measuring the values of the predictor variables only when the behavior occurs. Regardless of the study methodology, the use of general linear mixed models (GLMMs) for statistical analysis permits stronger inference about the effects of predictor variables than is provided by older statistical methods. GLMMs allow researchers to disentangle from the error term the variation that is due to consistent differences between focal animals, groups, years, food species, or other uncontrolled but repeatedly sampled categorical variables that affect our observations. Reducing the error term then allows more powerful inferential tests of the effects of the hypothesized causal variables. All field studies will benefit from a better quantitative theoretical framework within which to interpret the match between expected and observed outcomes.

**Keywords** Control · Experiments · Field studies · Observations · Primates · Repeated observations · Statistics

---

C. H. Janson (✉)  
Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA  
e-mail: Charles.janson@mso.umt.edu

## Introduction

Despite their common roots in natural history, field studies of ecology and primatology have diverged in methods. For several decades ecology has adopted a strong experimental approach to test increasingly refined mechanistic explanations, usually with large sample sizes. Field primatology remains a generally observational discipline focused on relatively small numbers of individuals or groups, with proportionately few field experimental studies, most of which focus on vocalization playbacks. This difference between disciplines does not derive from a lack of appreciation of the power of experiments by primatologists because experiments have been common in psychological studies on captive primates dating back  $\geq 50$  yr (to the early studies of Harry Harlow and colleagues). Rather, field studies of primates have seen limited use of experiments for 2 main reasons: 1) There has been an overriding emphasis to describe and emphasize the diversity in the natural behaviors and ecology of primate species; and 2) field primatology emerged largely from a descriptive anthropological tradition, rather than a quantitative biological or psychological one. As a result of both of these traits, students have, until recently, been encouraged to describe the basic ecology and behavior of novel populations or species rather than delving more deeply into the biology of already well-described populations. In principle, such broad sampling should have led to robust conclusions about the mechanisms operating to generate repeatable patterns across many natural populations, as has happened in some areas of ecology (Sagarin and Pauchard 2010). In practice, the lack of a mathematical theoretical framework that focuses attention on precisely defined variables, the coexistence of distinct conceptual foci (biological, psychological, and anthropological), and the repeated emergence of newer technologies, e.g., for genetic or nutritional analysis, have led to a large literature with results that are only occasionally directly comparable between species or populations.

As global-scale alterations of habitat and energy use by humans generate increasingly frequent and acute management challenges for natural populations, the public (through national funding agencies for scientific research) have sought more broadly applicable answers from scientists about the causes and possible solutions to environmental problems. This has led to recent reviews in the ecological literature that reflect a rekindled interest in the value of observational science and the limitations of the experimental method in ecology (Hewitt *et al.* 2007; Pickett *et al.* 2007; Sagarin and Pauchard 2010). The main tension identified in these reviews is that between the strength of inference (what I call rigor) and the ability to extrapolate the results (what I call the range of applicability). In general, experiments excel in rigor, but have limited range, whereas observations are considered less rigorous in inferring mechanisms but have a range that encompasses at least the full spectrum of natural variation (Sagarin and Pauchard 2010). In simple systems, including the extreme ideal in which only 2 mutually exclusive hypotheses can be tested against each other (strong inference: Platt 1964), experimental approaches can yield results with both good rigor and broad range (Hewitt *et al.* 2007). However, in complex systems, such as those typical of primate biology, experiments are difficult and limited in scope, and so their conclusions may be difficult to generalize. Experiments can also extend our ability to observe phenomena beyond conditions present in

natural conditions, such as asking whether a species is capable of thriving in a location where it does not currently occur. All these reviews call for a pluralistic approach, either explicitly embedding experiments in the context of observational studies, or using an iterative framework in which experiments and observations are used to refine each other within the guidance of a conceptual or mathematical theory.

I here address many of the themes of these methodological reviews with reference to primate behavioral ecology and my study system. First, I present some of the logical and statistical difficulties of observational studies in inferring the biological importance of several putative causes, with some recommendations to avoid a few of the inherent problems. One solution to strengthen inference about causes is to perform experiments, but I argue that experiments and observations are not strict alternatives but endpoints of a continuum of control imposed by the researcher. Then I describe some of the practical limits of observational research, followed by a section on its benefits including breadth of data obtained and openness to detecting novel patterns. Turning to field experiments, I briefly review the successful projects at our study area in Iguazú National Park, Argentina (25°40.7'S, 54°27'W), in which experiments have contributed rigor and insight into processes not otherwise easy to observe, such as food-related cognition or the detection of predators by monkeys. As a cautionary tale, I then give a detailed example of practical limitations and outright mistakes in developing a set of social foraging experiments using feeding platforms. This example emphasizes the importance of embedding experiments within a thorough understanding of organismal natural history available only from observational studies. In keeping with the proposed gradient between observations and experiments, I introduce the middle ground of quasi-experiments that may combine the best of both methodologies. An additional useful tool, statistical analysis using general linear mixed models, improves on other statistical methods by providing a better separation between the effects of putative causal variables and those of uncontrolled but repeatable categories such as focal animal identity. I end with a plea for more extensive use of quantitative modeling to sharpen the predictions used in primate social and ecological research, as one way to increase the power to reject or accept particular hypotheses. Because I present no detailed data, I structure the paper as an extended review and discussion instead of the traditional sections of Methods, Results, and Discussion.

### **Limitations on Inference from Experimental vs. Observational Studies in Field Primatology**

Experiments are often held up as the true standard of science (Pickett *et al.* 2007, p. 197), the ultimate test of a postulated mechanism thought to produce a particular effect or pattern in the real world. A general definition of experiment is given in Pickett *et al.* (2007, p. 48): “Manipulation of a system to generate a reference state or dynamic of known characteristics.” Experiments are excellent procedures to determine causes when these are few in number and mutually exclusive, as enshrined in the postulates of strong inference (Platt 1964). Unfortunately, in the study of complex ecological systems or complex behaviors of animals, the possible number of explanations for a given pattern or behavior is often large, the explanations are rarely mutually exclusive, and the predictions emerging

from each explanation often overlap. For example, various authors have postulated 4–14 distinct hypotheses to explain food-associated calls given by ravens and monkeys (Di Bitetti 2001; Hauser and Marler 1993; Heinrich and Marzluff 1991). A selection of predicted consequences of a subset of these hypotheses is given in Table I. The main point to note about Table I is that for each context or consequence (column), the various hypotheses (rows) make different predictions about whether the column variable should increase, decrease, or not change depending on whether an animal gives food-associated calls. Because the hypotheses are not mutually exclusive,  $\geq 2$  of the hypotheses could be in fact be operating at the same time. If the effects from different hypotheses are additive then the overall observed effect could be positive, negative, or indistinguishable from 0, even (or especially) if several of the hypotheses are actually correct. For instance, if food-associated calling serves to announce food possession but also is favored by kin selection, then it would be possible for the combined effect of both hypotheses to yield no notable difference between callers and noncallers in their distance to neighbors, feeding rate, reproductive success, and the likelihood of the caller's being male or female, leaving only a predicted difference in social status. This difference in social status would correctly indicate the importance of food possession, but the effect of kin selection would be missed entirely. Only in the simple case that exactly 1 explanation is in fact true could there be clear support for 1 of the hypotheses vs. the remaining ones.

If just about any observed outcome is consistent with the additive effects of some set of hypotheses, then how can one hope to decipher the actual causes affecting a phenomenon of interest? One alternative would be to perform a series of

**Table I** Predicted differences between individuals giving vs. not giving food-associated calls

Hypothesis	Prediction				
	Distance of neighbors to caller	Caller's food intake	Caller's mating success	Caller's sex	Caller's social status
Kin selection	Decrease	Decrease	No change	Female > male	No effect
Group selection	Decrease	Decrease	No change	Female > male	No effect
Enhance exploitation efficiency	Decrease or increase	Increase	No change	No sex difference	No effect
Individual food defense	Increase	Increase	No change	Male > female	High > low
Attract allies for food defense	Decrease	Increase	No change	Female > male	Low > high
Reduce predation risk	Decrease	Decrease	No change	No sex difference	No effect
Mate attraction	Decrease or increase	Decrease	Increase	Male > female	Low > high
Enhance social status	Decrease	Decrease	Increase	Male > female	Status increases

Simplified from Di Bitetti (2001, pp. 245–246). In cases where different directions of responses are predicted for a single cell, the different responses depend on the caller's social status

experiments in which the strength of each hypothetical cause is varied independently of the remaining hypotheses and the resulting changes in predicted effects are documented. An experiment might not be possible for all hypotheses, but at least some hypotheses might be supported or rejected with confidence. A similar approach could be taken by stratifying (selecting) unmanipulated observations according to the strength of the putative causal hypotheses, e.g., predation risk, and then analyzing the relationships of the predicted effects to the levels of the hypothetical cause. Again, not all possible causes might be amenable to such analysis either because they do not vary, or such variation is not easily defined, e.g., the strength of group selection, or identified, e.g., different kinship structures.

The main drawback to the observational approach is that levels of 1 potential causal factor, e.g., predation risk, may covary with other possible causal factors, e.g., food availability, mating season, group size, and kinship structure, so that the pure effect of variation in a given causal factor cannot be measured in isolation. This problem can, in theory, be solved with multiple regression to “control” (or more properly, account for) the variation in other possible causal variables so as to isolate the independent contribution of the variable of interest. Though I am far from an expert in the underlying theory of statistics (Sokal and Rohlf 1995), my experience as a critical end-user is that multiple regression can be finicky. First there is the problem of collinearity: several variables may so highly correlated with each other that their effects cannot be isolated in practice, or the results will depend very sensitively on the particular data points or other variables that are included in the analysis. There are methods to circumvent this problem, e.g., using synthetic uncorrelated axes derived from principal components analysis, use of ridge regression, and use of information criteria to perform model selection while accepting that several models may be indistinguishable in terms of explanatory power, but each method carries additional assumptions or complicates interpretation. Second is the issue of nonlinearity. Because parametric regression and other variations of general linear models are designed to detect only linear relationships, if any of the relationships between the predictor variables and the response variable or among the predictor variables is not linear, e.g., follows a curved or “humped” relationship, then the strengths and values of inferential tests of all the hypotheses are suspect. It is possible to allow for some kinds of nonlinear relationships among the variables by suitable data transformations (Sokal and Rohlf 1995), but these are often cumbersome and can result in analyses with variables that are difficult to describe and understand biologically. Third, there is the problem of outlying data points: those with values much higher or lower than the rest of the data. If any of the predictor variables includes a small number of such outliers, then the strength and slope of the relationship of that variable to the response variable may be much higher than if there were no outliers. It can even occur that 1 or 2 extreme outliers can “create” a significant relationship that does not exist in the main cluster of data points. Many diagnostic methods exist to identify such influential data points (Bollen and Jackman 1990), but they do not provide a simple way to decide how to deal with them; if these points are valid, they should not be eliminated from the analysis arbitrarily.

A simple solution that often works to resolve the second and third problems is using the log-transformation of the original data (it does not matter if you use natural

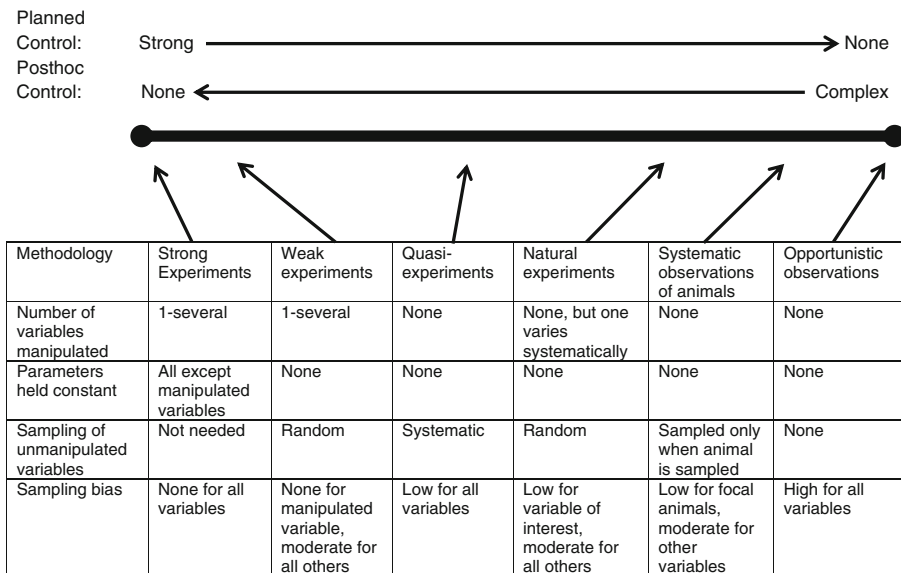
or base-10 logarithms, as they differ by only a constant). Two log-transformed variables that have a linear relationship with a slope different from 1.0 can have a variety of curved relationships in the original (untransformed) scale: decelerating (slope between 0 and 1), accelerating (slope greater than 1), hyperbolic (slope of  $-1$ ), etc. These allometric relationships are common in biology (Gould 1966). Log-transformations of the original data also make high outliers fall much closer to the remainder of the data, thereby making the regression slopes much less sensitive to these outliers. Log transformations are not a panacea, and using them can create other problems; for instance, small outliers can be moved much further from the remaining data after the log transformation, although this is easily solved by adding a constant, such as 1.0, to all the data before applying the log transformation.

In general, it is dangerous to trust the results of any multiple regression analysis in which the pairwise relationships among the data are not easy to visualize. The ability of the researcher to visualize the relationship is essential to check for the 2 main assumptions underlying parametric regression: 1) normality of the residuals from the regression (not normality of the raw data for either the response or the predictor variables, no matter how many published papers incorrectly assert this criterion) and 2) homoscedasticity (the uniformity of the scatter of the residuals, the variance of which should be independent of the values of the predictor variables). Although it is technically feasible to study the relationships among all pairs of variables even for large numbers of variables, it is my experience that once multiple regressions include more than *ca.* 5 predictor variables, the results tend to be unstable to small changes in included data and variables. This instability may result because subtle nonlinear relationships among several variables can fail to be detected, yet they may have marked net effects on the magnitude of the residuals of the analysis if the relationships are incorrectly modeled as straight lines.

The best way to reduce the complexity of the problem when analyzing multiple regressions is to use the power of theory to narrow the universe of likely predictors and confounding variables to a manageable number. For instance, when trying to understand the relationship between individual food intake and group size in wild primates, it is essential to consider the confounding effects of patch size on both food intake and feeding group size (Janson 1988b), and it is necessary to realize that the relationships may differ between food species because of other unmeasured but potentially important variables such as patch density or food nutritional content (Janson 1988b). These may be termed causal or control variables. However, there are many other potentially confounding factors that have no theoretical reason to correlate with the main variables of interest; such “noise” variables might include the season of the observation, different “personalities” among sampled social groups, and the year of the study. For these noise variables, an exploratory analysis should be used to reveal if they have any notable relationship to the response or main predictor variables; if they do not, there is no logical reason to retain them in the analysis (but their exclusion should be made clear in the study’s Results). I use the word notable rather than statistically significant on purpose because the latter depends on sample size, but logically I would want to consider any confounding variable that explained a notable fraction (I use a minimum of 25%) of the variation in either the predictor or response variable. Modern statistical treatments allow for sophisticated model selection according to a variety of criteria, e.g., AIC (Akaike

information criterion: Bozdogan 1987), and these are generally to be preferred over older stepwise models that use simple statistical significance as an all-or-none criterion to include a variable in the analysis (Mundry and Nunn 2009).

In the end, the purpose of either experiments or multiple regression analysis of unmanipulated observations is to control variation in  $\geq 1$  of the predictor variables, so as to reveal more clearly the effect of each variable on the response variable. Although often described as strict alternatives, experiments and observations merely represent ends of a continuum of planned vs. *post hoc* control (Fig. 1), and there are intermediate points on this continuum. Even within experiments, there are differences in design, from the extreme control implied in strong inference (Platt 1964), in which one variable is manipulated and all other nontested variables are excluded or held constant, to field experiments, in which 1 or a few factors are manipulated and all other variables change in unknown ways over time and space. Unmanipulated observations run from natural experiments, wherein 1 or a few variables vary in a known systematic way, e.g., along an environmental gradient or through time, and the other variables are not known, to opportunistic data wherein many variables may be measured but none are controlled or vary systematically. In between manipulative studies (experiments) and unmanipulated studies (observations) lie what I call quasi-experiments, a realm of focused observations taken under conditions that account for variation in 1 or a few hypothesized causal variables, although without any actual manipulation of those variables. Because all of these



**Fig. 1** The continuum of planned vs. *post hoc* control of variation in predictor variables in scientific studies. Various kinds of experiments cluster at the more planned end, while opportunistic observations are at the extreme of unplanned studies, which require considerable *post hoc* analysis to allow some confidence in inferring the causes of observed patterns in a response variable of interest. In between these 2 extremes, quasi-experiments use highly systematic observation of a wide range of variation in the predictor variables to offer more robust inferences about causality than is allowed by opportunistic or even systematic observations on focal individuals.

methods share the purpose of revealing causes underlying observed patterns, 1 of them is not universally better or worse than the other, but each can be applied under different circumstances and with different limitations (Table II).

As recommended by Hewitt *et al.* (2007), experiments should ideally be embedded in the context of observational data. In practice, this means that an experiment designed without knowing the relevant biological constraints on the organism (such as its ability to perceive or avoid the experiment) risks being at best a waste of time and at worst misleading or even harmful to the study animals. Thus, one important reason that experiments have been scarce in primatology is that few populations have been studied for long enough that researchers know the relevant constraints. In addition, relatively short periods of study have meant that the rate of new discoveries about behavior have outpaced the ability to absorb and integrate this information into a conceptual framework that allows strong predictions to be made. One of the benefits of long-term studies of particular primate populations should be the ability to design experiments that are both informative about mechanisms underlying observed behavioral patterns and ethically justifiable within the context of studies on long-lived individuals in often-small populations. Although the roster of primate species that have been the subjects of long-term ( $\geq 20$  yr) studies is still relatively small, it is growing rapidly (Kappeler and Watts 2012). Even if experiments might not be feasible in all of these populations, information from these long-term studies could be used to design and implement experimental studies in less thoroughly studied populations of similar species elsewhere.

## Practical Limitations and Benefits of Observational Studies

### Limitations

It is nearly trivial, but important, to point out that there is no such thing as unbiased observation. Every human being, or even a team of observers, cannot attend to every

**Table II** Expected effects of the use of different study methodologies (columns) on rigor and range of outcomes (rows)

Criterion	Observations	Quasi-experiments	Experiments
Feasibility (range of situations to which method can be applied)	High	High	Low/intermediate
Range of parameters observable	Natural range	Natural range	Can be more extreme, or more constrained than natural variation
Data yield	High	Intermediate	Low
Ability to document non-events	Low	High	High
Ease of interpretation	Low	Intermediate/high	High
Openness to novel discovery	High	Low/intermediate	Low



possible behavior, state, and context for even a single animal, let alone many members of a social group. Thus, an important part of any field study is deciding what data you are not going to take. If you are interested in social behavior, you are not likely to be able to measure details of ecology, and vice versa. If you are interested in the relationship between ecology and social behavior, you will need to decide how to compromise between them. In the early comparative studies of New World primates conducted in Manu National Park in Peru (11°53.3' S, 71°24.45'W; Terborgh 1983), we deliberately focused on feeding ecology as the main subject of study, recording ecological behavior and contexts in considerable detail. In contrast, we recorded all social behaviors *ad libitum*, and essentially did not record some (such as vocalizations) at all. These data produced detailed descriptions of food choice, patch size, travel behavior, and group size (Terborgh 1983), but left most aspects of aggressive and cooperative social behavior for later studies.

In my thesis work, I focused on some of these details for only 1 pair of capuchin species at Manu, integrating ecological context and ranging behavior with systematic observation of aggressive behavior, grooming, spatial position within the group, and other aspects of social structure. Lacking a broad theoretical framework to guide my research (I started in 1973, only 2 yr after the seminal papers on the benefits of sociality in reducing predation risk: Hamilton 1971; Vine 1971), I was forced to be a generalist, collecting and describing every plant species the monkeys ate or could have eaten, recording individual activities in second-by-second detail so that I could use them to develop crude energy budgets, labeling every important tree that members of the group visited even once, and monitoring the phenological status of hundreds of trees along 8 km of trail. John Terborgh once accused me of taking data on everything, without hypotheses, but that was far from the case; I was just interested in everything at once!

There were several disadvantages of my generalized observational approach. First was simply fatigue. Because there was vastly more to describe than I could possibly accomplish, I filled nearly every waking minute with data collection of some kind. I was convinced that otherwise I might miss some contextual data that would be important in figuring out the behavioral patterns that characterized these 2 species of capuchins. Second was the fundamental inability to describe certain parameters of interest. Some were important variables that were difficult to infer directly from observations, such as which food sources were retained in an individual's spatial memory, whereas others were interactions that were simply rare and hard to observe, such as ability of predators and prey to detect each other in a closed rain forest environment. Third, the conceptual model that eventually emerged from this research (Janson 1988a) was a *post hoc* summary of observations, not a theory derived from first principles. Thus, it was vulnerable to the criticism that it was "only" a theory and that it might confound or miss entirely the true causal variables responsible for creating the patterns on which the model is built. The only way to counter such an argument was to manipulate the critical variables in the system so that the likelihood of chance correlations between these observed variables and possible hidden causal variables was very low, if not 0. To achieve this goal I followed traditional practice by using experiments, but first I need to point out the benefits of observational studies.

## Benefits

Observational studies can be quite efficient in generating scientific knowledge, allowing a researcher to address new questions and revisit old ones from a single original data set. The end result of my dissertation research was a mass of data derived from 48 mo of observation, a treasure trove that allowed me to describe quite a variety of basic patterns of socio-ecology. These started with the main focus of my study: aggressive competition and its effects in structuring food competition and spatial positions of individuals (Janson 1985, 1990a, b), but ranged to a detailed comparison of social structure between 2 superficially similar but in some ways profoundly different capuchin species (Janson 1986), the mechanisms and magnitude of scramble competition in 1 species (Janson 1988b), and the adaptations of some plant species to monkeys as dispersers (Janson 1983). Inspired by these results and incorporating those of many other field studies in the 1970s and early 1980s (Janson 1988a, b), a coherent explanation of the patterns began to take shape. The general conceptual theory for scramble competition (Janson 1988a) can be summarized in the phrase “primate social groups are economical foraging machines.” In such machines, the “parts” are the foraging group members and the “yield per part” (net energy intake) is a relatively simple function of the rate at which the “machine” encounters food, the value of the food it encounters, and how many parts share or divide the encountered food. There are interesting differences among species in constraints that affect these values, such as the digestive capacity of each individual, how far away a primate group can detect novel food sources, how predation pressure favors spacing between foraging group members, what foods dictate the minimum foraging effort of groups, and how memory of renewing food patches affects foraging strategies and success. Several ensuing papers have dealt with these and other similar issues (Janson 1998; Janson and DiBitetti 1997; Janson and Goldsmith 1995; Janson and Vogel 2006). Ultimately, the diverse observations I made during my dissertation research directly supported the writing of 14 data papers and at least as many more conceptual and, eventually, experimental works. I am still mining these data today, >30 yr later!

An additional benefit of observational studies is their flexibility. In particular, I was open to noting and systematically studying any behaviors of interest that emerged during the field work. For instance, we were struck by observations of a behavior so unusual that for several years, we thought that female brown capuchins were periodically stricken by some illness causing such physical distress that they sought the company of the group’s dominant male for comfort: they moaned loudly, grimaced as though in pain, clutched their bellies, persistently followed the male, and attempted to touch him, sometimes for several days on end. Finally, sometime in the third year of study, we happened to have observation conditions good enough to see the “sick” female mate with the male (an act that occupied only about 1 min out of every day), and at least for a while the other behaviors stopped. This discovery of an extremely active female role in mate choice and mating behavior was totally novel to us, raised as we were on 1960s descriptions of Old World savanna primates, with male challenges over estrous females and the female apparently passively accepting to mate with the victorious male (Hall and deVore 1965; Hausfater 1975). More systematic study of this discovery led to a thorough description of what was

then quite a novel mating system (Janson 1984), although it required a change in methodology: as soon as we noted the presence of an estrous female capuchin, we deliberately made her the focal animal for extended periods, disregarding the other data collection for a while. Thus, with modest compromises, we were able to detect and describe interesting novel behaviors, even if they were not the primary focus of our original study.

### Practical Benefits and Limitations of Field Experiments in Primatology

I had decided as early as 1980 that it would be useful to perform experiments in the field to test certain hypotheses, particularly ones concerning the mechanisms apparently underlying observable patterns. The bulk of the successes of this approach are documented in various papers by my students, collaborators, and me, and I only briefly summarize them here. Less publicized are the difficulties and indeed outright failures of field experiments, which I will recount in greater detail to allow others to avoid my mistakes.

#### Successes

At our study site in Iguazú, Argentina, my colleagues and I have used experiments in the field to manipulate and test the importance of a variety of mechanisms underlying observable behaviors of capuchins and coatimundis, including ranging behavior (Hirsch 2010; Janson 1996; Janson and DiBitetti 1997), spatial cognition (Janson 1998, 2007a), predator recognition (Janson 2007b; Wheeler 2010b), intergroup competition (Scarry, unpubl. data), intragroup contest competition (Janson 1996), functions of vocalizations (Di Bitetti 2003, 2005; Wheeler 2008, 2010b), deceptive communication (Wheeler 2009, 2010a), and inherent interest in or avoidance of novel objects (Visalberghi *et al.* 2003). Common to all of these successful studies were several important features: 1) the focal individuals could either not avoid the experiments, e.g., vocalization playbacks, or could participate in the experiments without much compromise to other activities; 2) at least in food-manipulation experiments, the subjects were attracted to participate in the experiments in part because in the subtropical winter (June–August) little fruit has usually ripened and thus the food provisioning (bananas or tangerines from local markets) was a highly valued supplement to their diet; and 3) enough descriptive natural history or preliminary study on this population existed to allow reasonable guesses about how to set up the experiments and what questions could realistically be addressed.

The power of doing field experiments at our site lay in 2 distinct arenas. First, by deliberate design of parameter combinations, we could create essentially uncorrelated variables across treatments, thereby making interpretation of the emerging results much easier. For instance, in experiments to manipulate the intensity of intragroup contest competition, it was possible to set up feeding sites that provided distinct, controlled combinations of food average amount, variance in amount, and spacing (Janson 1996). Second, the experiments could sometimes expand the range of parameters or phenomena that were available for observation to realms rarely available in natural circumstances. Examining behavior in these realms is not

necessarily “unnatural,” although a researcher should always be humble in interpreting the outcome of an experiment, aware that the experiment, however elegant in design, may not test the hypothesis proposed. Perhaps the best examples of experiments outside the “natural” box are those exposing capuchin monkeys to models of predators (Janson 2007b; Wheeler 2008, 2010b). In return for an arguable small loss of realism, these experiments allowed researchers to define the context and behaviors of the focal animals before the detection of a “predator” because we knew in advance approximately where such detections must occur. This method has provided insights into how and why prey detect predators, in ways that simply had not emerged from a prior half century of field studies of primates in natural circumstances. Equally important, these experiments allowed us to observe those situations in which the capuchins did not detect “predators”; non detection of natural predators by monkeys is almost never documented by human observers unless the predators are themselves the focus of study (Zuberbuhler *et al.* 1999). Analysis of both detections and non-detections revealed the unexpected fact that capuchins are amazingly poor at detecting camouflaged nonmoving predators: a capuchin group of >20 individuals passing directly over an ocelot model placed at random (neither particularly concealed nor open) on the forest floor still had a >50% chance of not detecting the model at all (Janson 2007b)!

### Failures

There were several reasons that some field experiments failed in my studies. First, the focal individuals were not always sufficiently interested in the experiment to participate. This occurred primarily because of neophobia. In the first case, during my dissertation research in Manu National Park, I had hoped to use bananas cultivated near the Park to manipulate food abundance and distribution as I eventually did in Argentina (Janson 2007a). However, bananas were a novel food source and, despite their superficial similarity to a local fruit (*Jacaratia digitata*), few individuals ever tried them during my 18-mo dissertation project. Maddeningly, the following year, when I was not able to return to the field site, the main study group decided that bananas were choice edibles, and the capuchins became an actual annoyance, removing banana pieces from small-mammal traps and destroying bunches of bananas stored at the main research building. In the second case, I was using the platforms that were a common fixture of experimental work at Iguazú, but with a study group that had never experienced them before. Even though other study groups had learned to overcome their fear of the platforms to acquire the fruit contained in them, this study group resolutely ignored the platforms, no matter how much food was there when they passed by. Only when Celia Baldovino cleverly camouflaged the platforms with branches from local naturalized citrus plants (with attached citrus fruits) did the monkeys approach the platforms and stop to feed, initially in the attached branches and later in the platforms themselves. Without such field insight and persistent effort, this experiment would never have continued.

The second reason for experiments to fail was the inability to design the experiment to fit within the lifestyle of the focal individuals. This led to the most comical (if frustrating) experimental failure of my career. I had proposed to study foraging choices in the Iguazú capuchins by setting up the equivalent of a Y

maze, a standard and powerful experimental tool in behavioral or psychological studies of captive monkeys. There are 2 important features of such a maze: 1) the subject starts at the same point each time, and 2) the subject is then allowed a choice of 2 alternatives, but can experience the outcome of only 1 alternative per trial. In my naiveté, I believed that I could get the main capuchin focal group to use the same “favorite” sleeping area on many successive nights, so that they would begin each day’s foraging at the same start site. I began by trying to get 1 group to return to feed in the late afternoon each day at a platform site near their most frequently used sleeping area, hoping that they would then stay in the vicinity and use that sleeping area. Knowing that the group could easily travel 200 m/h, it was important to have the group feed at this platform late in the afternoon, ideally after 16:00 h. The plan was to reinforce afternoon feeding, but not morning feeding, gradually narrowing the criterion time period during which the group would receive food. At first, this seemed to work well, as the group discovered and visited the site on several afternoons in a row. However, when we started to tighten the criteria by feeding only at visits in the later afternoon, the paradigm fell apart. After the first few times that we did not feed them in the afternoon (because they visited too early), the group left and returned the next morning, when they again did not get fed. Rather than guessing that they would be fed later the same day (as was obvious to us), the capuchins apparently assumed that this platform “tree” had run out of fruit so they stopped coming to the site altogether unless they happened to be in the vicinity anyway. In any case, we completely failed to mold their behavior to the purpose of consistent use of 1 sleeping area. In retrospect, we were probably asking them to violate 3 distinct “rules” of wild ranging behavior: 1) do not sleep in the same area on consecutive nights (Di Bitetti *et al.* 2000); 2) trees with large fruit do not ripen fruit on any conspicuous daily schedule so there is no reason to return to a particular tree at any given time of day; and 3) trees that produce no fruit at all for a day and have no evidence of green fruit available are most likely exhausted and so are unlikely to provide more fruit in the future.

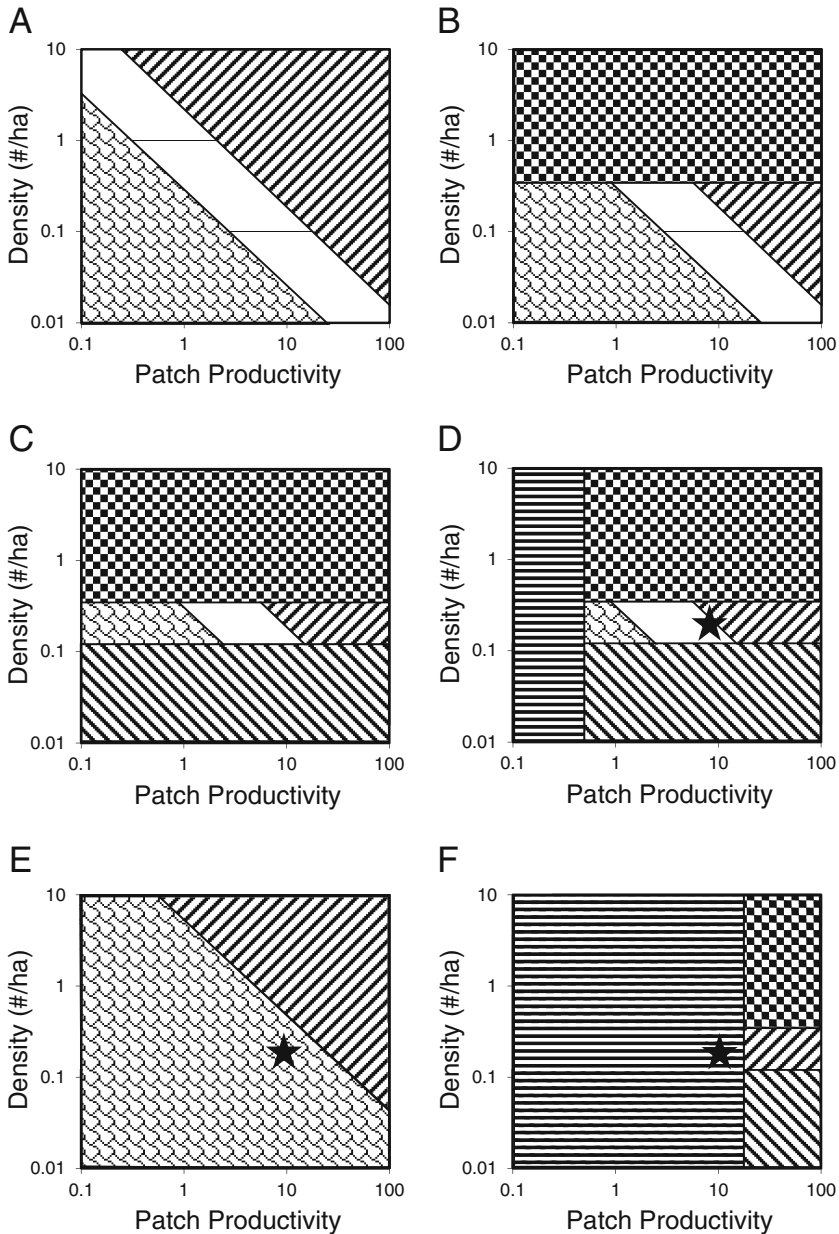
Having failed to train the group to use a single sleeping site, I decide to try a more direct “conditioned response” method to train them to use a single starting site to feed. We had a cow’s bell that we set up at the start site. As the group approached the feeding area, we rang the bell as we put food on the platforms. Very quickly, the animals learned to associate the bell with the presence of food at the starting site and would come running quickly to those platforms when we rang the bell. This seemed like a victory for classical training methods, but brought with it a different problem. This occurred because the group would often pass by some of the other feeding sites (“goals”) on their way to the starting site. By design (the purpose being to simulate a Y maze), we did not feed them at goals when they visited them before the starting site. However, the group reacted to the lack of food in goals as they would to a fruit tree with no fruit: They avoided going back to the goal that day. To try to convince them they could receive food at goals even after finding them empty, we would blow a whistle when the group was near a goal site, and then reward them with food if they visited it. Not surprisingly, in retrospect, the monkeys learned that “whistle means food” and by extension “no whistle means no food.” Soon the group refused to go visit any goal, even if quite close, unless we blew the whistle first.

**Fig. 2** Design constraints on an experimental array of platform feeding sites. **a** The area in the upper right corner denotes the fact that if the total productivity (= patch density times patch productivity) of the array is too high, the group will visit only a portion of the array each day, whereas the area in the lower left corner reflects the minimum total productivity of the experiment needed to entrain the group's foraging effort in the face of normal winter fruit production. **b** The area added to the top of the graph reflects the constraint that the feeding sites must be  $\geq 200$  m apart (mean density of 0.25/ha) so that the group does not visit  $>1$  site at a time. **c** The area at the bottom of the graph restricts the 5-site array so that it fits within the largest single block of relatively flat homogeneous habitat, ca. 50 ha. **d** The area on the left margin of the graph reflects the fact that the group will not visit individual feeding sites that are too unproductive relative to natural food patches available in winter. The unshaded area in the center of the graph denotes the small universe of possible combinations of feeding site density and productivity feasible for the experimental goals. The star denotes the most desirable combination of parameters for a feasible experiment. **e** When the winter total natural fruit productivity increases (during warmer winters), the feasible feeding array (star) as a whole is not productive enough relative to the natural productivity level to keep the group's attention focused solely on the experiment. **f** When a particularly productive individual fig tree happens to fruit during the experiment, the minimum threshold of feeding site productivity increases dramatically, making the feeding sites of a feasible experiment not attractive enough to entrain the group's foraging movements.

Inadvertently we had succeeded in training the capuchins to use only human cues for foraging, thus completely defeating the purpose of the study!

Once we realized our error, we quickly gave up on the training idea and redesigned the experiments so that, no matter which of the feeding sites the capuchins used first in the day, they would then always be faced with a choice of traveling to a closer, less-rewarding goal vs. a more distant, more rewarding goal (Janson 2007a). In reality, as with natural food patches, these were not mutually exclusive choices, because we had learned that we had to provide the monkeys with a predictable reward for visiting a feeding site, regardless of the order in which they did so. Analysis of the sequences of site choices quickly revealed that the groups did not choose among sites by comparing each site against the others, but integrated the expected rewards and effort across sequences of sites (Janson 2007a).

Given that capuchin groups appear able to integrate rewards and distances across  $\geq 2$  feeding sites, it seemed interesting to test whether capuchin groups could solve simple traveling salesman problems (TSP) involving several sites, as macaques appear to be able to (Cramer and Gallistel 1997). Having learned from the Y-maze experience, I designed arrays of 5 feeding sites that revealed something about the possible rules used to guide their foraging routes, regardless of which site the monkey group chose to use first on a given day (Janson 2000b, p.196). The details of the experiments and their results will be presented elsewhere, but the focus of interest here is how we designed the overall placement of the feeding sites, and how much food to put on them, to achieve the experimental goals. This turned out to be quite a complex decision, because the array was most informative only when all the sites were visited in a single day, preferably in a single sequence with no other (natural) feeding sites interposed between the experimental sites. So, the sites had to be attractive enough to keep the monkeys interested in using them to the (near) exclusion of other natural foods, yet could not be so rewarding that the monkeys visited only a subset of the sites each day (Fig. 2a). Further, the sites could not be too close together, or else the group would split apart to use multiple sites at the same time (Fig. 2b). However, if they were too far apart (Fig. 2c) the shortest travel routes between some pairs of sites would cut across  $\geq 2$  very distinct habitat types. The different costs of movement in various



habitat types would complicate any attempt to use distance (easily measured) as a proximate variable for travel cost (which is the true currency in TSP, but is hard to measure). Finally, each site had to be minimally rewarding compared to natural alternatives if the group were going to prefer visiting the platform sites first (Fig. 2d). The final design parameters were chosen from the rather small “space” of possible values delimited by the previous constraints (Fig. 2d).

It turned out that we could conduct these experiments only when the subtropical winter was cold enough to make the background availability of competing feeding

sites low and unrewarding, so that our feeding sites were very attractive to the monkeys. In warmer winters (increasingly frequent starting after 1999), the group would either mix feeding platforms with natural food sources, or would not visit all the platform sites each day (Fig. 2e). A similar problem occurred even in cold winters if a particularly productive fig tree ripened fruit then (Fig. 2f), as the group would then ignore the platform sites or visit them only after gorging at the fig tree.

In sum, field experiments can be useful to 1) disentangle complex sets of correlated variables; 2) extend the range of parameters that are observable in a field setting, including allowing “non-events” to be recorded; and 3) set up specific situations that conform (to a certain degree) to the assumptions of particular theories. At the same time, field experiments may be tightly constrained by the background levels of competing stimuli (food in the case of feeding experiments, noise, or incompatible social stimuli in the case of vocalization playbacks). Designing experiments that attract the participation of wild animals without fundamentally changing their biology is a challenge and almost an art, in the sense that it is necessary to make decisions that ultimately rely on an aesthetic or intuitive sense of how your focal species will react to the experiment, with only rough quantitative constraints as guides.

### **The Middle Ground: Quasi-experiments and Random-Effects Statistics**

There is a middle road between opportunistic observations and highly constrained experiments, represented by planned observations focused on one or a few specific questions (Fig. 1). There is a long tradition of such systematic observation, focused on behavior, dating back to the seminal methods paper of Altmann (1974). This paper introduced the notion of systematically sampling the behavior of focal individuals rather than opportunistic sampling of any individual or behavior that was visible or happened to be doing something “interesting.” Such systematic sampling both allows more confident statements about variation between individuals and also reduces the sampling bias inherent in opportunistic methods. Two developments have increased the rigor and range of these methods: 1) extending systematic observation to the ecological or social contexts hypothesized to be the causal links explaining the behavior’s occurrence and 2) the more widespread use of random-effects statistical models to make maximal inferential use of the repeated sampling of focal animals that is common in primatology (and indeed can be considered one of its strengths).

Good recent examples of systematic sampling of contexts are the foraging cognition studies of Janmaat (Janmaat *et al.* 2006a, b). These studies use strong hypotheses to set up a specific sampling design that allows more complete or rigorous testing of the hypothesis, at the expense of gathering other opportunistic data. Rather than following a group of monkeys and recording all the trees that they visited, Janmaat preselected a set of focal trees (the ecological context) and recorded what happened whenever a group approached within 100 m of any focal tree. This focus on tree choices rather than on monkeys meant that it was possible to contrast systematically the traits of trees visited vs. those not visited under comparable conditions. This approach allowed strong tests of the monkeys’ memory of such



variables as tree fruiting state and type of food (Janmaat *et al.* 2006a) or ripeness and expected crop size (Janmaat *et al.* 2006b). In particular, the ability to score non-visits to particular focal trees is similar to the ability to score non-detection of predator models in the experimental studies described in the preceding text.

In addition to quasi-experiments, random-effects statistical models are an important improvement to inferential tests of hypotheses in primatology. A recent excellent review of these methods is available (Bolker *et al.* 2009), but the essence of the method is to recognize that each focal animal (and other repeatedly measured categories, such as study year, group, etc.) is a block of data, within which observations are likely to be similar to each other, but different from observations in a different block. The causes of variation between blocks may not be known, hence the term random effect. The random effects model estimates the magnitude of the average deviation of each block's observations from the overall mean, assuming that the distribution of these block deviations is a normal curve. When applied to regression, such models can estimate random deviations among blocks for both the intercept and the slope of the regression; both should be explored.

There are several major benefits of these models in behavioral studies. By correctly stratifying observations by focal individual, these methods 1) account for repeated and (typically) uneven sampling on individuals when estimating the overall mean of the response variable; 2) allow for some autocorrelation among data sampled on the same individuals because these data all share a single common estimate of the random block effect; 3) allow explicit estimates of the random effect due to each level of the block, e.g., how individual X differs from individual Y, and whether 2007 was an exceptional year; and 4) provide an estimate of the overall contribution of the block type (focal individual, year, group) to the variation in the response variable. This last benefit is of great importance in inference, because it provides an estimate of the variation due to the block type, e.g., focal individual in general, not only for the particular individuals that happened to be used in this study. Other statistical treatments sometimes used in primatology, such as repeated-measures ANOVA of observations on individuals as fixed effects, also properly deal with the problem of repeated sampling, but limit the realm of inference to the specific set of individuals sampled. The simple solution of using only the average of the data per individual (or year, or group) also resolves the problem of repeated or uneven sampling, but it is very wasteful of data and does not permit an estimation of the specific contribution of interindividual variation to the overall pattern: individual variation and measurement error are confounded. Note that all of these statistical methods assume that sequential observations on the same individual are conditionally independent (given that you are sampling that individual); they do not correct for overly close sampling of the same individual through time, the problem of temporal autocorrelation (Janson 1990b).

The utility to behavioral researchers of random effects models has increased markedly in the past decade with the implementation of programs to allow their use with response variables that do not conform to the assumptions of conventional parametric statistics, i.e., homogeneous and normally distributed residuals from the fitted model. These generalized linear mixed models (GLMMs, mixed because they can include both fixed and random effects among the predictor variables) also allow correct estimation of model parameters when the response variable is a binary (yes–

no) outcome, a count variable (following a Poisson distribution), or other noncontinuous distributions that are common in behavioral data. A number of such GLMMs require specialized routines until recently available only in the statistical language R (<http://www.r-project.org/>), but they are starting to appear as options in common large statistical packages (SAS PROC GLIMMIX, SPSS MIXED procedure). As with any inferential statistics, it is important to make sure that the data conform to the assumptions of the method, but the complexity of GLMMs and the relative newness of procedures for their analysis require some additional caution on the user's part (Bolker *et al.* 2009); different programs may implement different procedures for estimating and testing components of the statistical model.

### The Importance of Theory in Inference

When testing any theory, there are 2 aspects worth examining. A theory is based on a particular set of assumptions, developed through mathematical or verbal logic to yield predicted patterns. The more commonly emphasized tests of theory concern the match between predicted patterns and observations (Pickett *et al.* 2007; Platt 1964). Less mentioned, but often easier to accomplish, is to make sure that the assumptions of the theory are met in the situation being studied. Verifying the assumptions of a theory is far more efficient because if an important assumption is not met, it cannot be the correct explanation for observed patterns, even if the theory correctly predicts them. Because, in a complex system, it is quite possible for a theory to predict observed patterns for the wrong reasons, testing the assumptions of the theory is an important prerequisite to accepting it as the explanation for observed patterns. Conversely, if the assumptions of a theory do not hold in a given system, then a mismatch between its predictions and observed patterns tells us little about its validity. For example, Grether *et al.* (1992) tested whether gibbons followed the predictions of the marginal value theorem (Stephens and Krebs 1986) when deciding to leave fruit feeding patches. The authors concluded that they did not, because gibbon feeding rates at the time of departure varied considerably among patches and varied only slightly within patches. Their rejection of the marginal value theorem as applied to gibbons might have been premature, because the predictions they tested depend tightly on a particular set of assumptions about the process of patch depletion. In particular, patch depletion in the conventional marginal value model is associated with gradually reduced food intake rates because of the time costs of searching for cryptic prey. Instead, fruit is typically conspicuous, reducing search costs and allowing intake rate to be dictated largely by handling time instead of searching time (Janson and van Schaik 1993, Fig. 5.2). In this case, intake rates are expected to remain constant during most of a feeding bout, declining only when nearly all fruits are removed, the very moment that primates would be predicted to leave the patch. Likewise, feeding rates at different patches would largely be dictated by differences in fruit handling time due to species or ripeness differences and should not be expected to converge among patches even just before leaving the patch. Incorporating more realistic assumptions about fruit patch depletion patterns into the marginal value theorem would have led to predictions that agree with the principal results of the observational study of Grether *et al.* (1992).

Testing the assumptions leading to a given prediction is especially important when the predicted patterns are given at only a coarse scale, e.g., increase vs. decrease, more or less likely, as is often the case in primatology (Table I). Using directional predictions means that if a given predictor variable has any consistent relationship with the response variable, it has a 50% chance of being in the predicted direction. Thus, it is quite possible for the theory to be upheld for a given variable, but in fact there is no causal relationship present. Such incorrect inferences are far less likely when the predictions of the theory are more precise, e.g., perfect scramble competition within a patch should lead to an exact inverse relationship between average per-individual food intake and the average number of individuals feeding (Janson 1988b). The use of more precise quantitative predictions in primatology would allow more convincing tests of postulated mechanisms as well as more frequent rejection of hypotheses that happen to be supported by occasional spurious correlations.

## Conclusion

Observations and experiments are not 2 distinct categories of methods for scientific study, but endpoints of a continuum of possible levels of planned control (Fig. 1). Recognizing this continuum reveals substantial variation within each category of methods, as well as suggesting the existence of intermediate methods (quasi-experiments) that provide some of the features of each extreme (Table II). With respect to primate field studies, the use of quasi-experiments has the strong advantage that it can be applied in many situations when manipulative experiments are either not feasible or might be considered unethical. Like experiments, quasi-experiments that systematically sample variation in the predictor variables, e.g., variation in fruiting state of focal trees (Janmaat *et al.* 2006a) can reveal situations in which a target behavior either occurs or does not occur. Knowing when a target behavior does not occur can be very revealing about the causes of the behavior when paired with detailed knowledge of the values of likely causal variables, whether such knowledge is derived from systematic sampling or from experimental control of the causal variables. Thus, the goal of experiments, to control large random variation in variables that might confound or mask the effects of truly causal variables, can now be partly achieved with strong hypotheses and structured observational techniques, combined with appropriately sophisticated statistical models. The only remaining strong benefit of experimental studies is the ability to design combinations of predictor variables that do not correlate with each other, thereby reducing the possibility that a particular predictor variable X is only by chance statistically correlated with the response variable, through the effects of some other “truly causal” variable correlated with X. Although well-designed experiments will always be more convincing than observations, it is no longer the case that observational studies need suffer from the stigma of being “natural history.”

The use of quasi-experiments can be strengthened even more by the use of recent advances in statistical modeling that permit the use of random-effects models with the kinds of data structures often found in behavioral studies: response variables with presence–absence outcomes, count data with discrete integer values, etc. Random-

effects models allow estimation of how uncontrolled but repeatable differences between measured blocks (focal animals, groups, years, etc.) contribute to variation in the response variable. By estimating the total variance contributed by each kind of block, the results should be robust to the choice of new or different blocks in future studies. Further, by including as block effects such variables as year or group, these analyses approach the level of detail in controlling extraneous variation that is the goal of manipulative experimental studies.

Regardless of methodology, the power to distinguish mechanisms or causes of primate behavior in the wild will benefit from closer attention to satisfying the assumptions of a particular model, and from more precise predictions leading to expected regression slopes, not just directional outcomes. Satisfying the assumptions of the theory that leads to the tested predictions can be an easy way to eliminate possible causes from consideration (Genty *et al.* 2009). More precise predictions of expected outcomes will allow researchers to refine what variables to measure and allow them to more frequently reject variables that happen to be correlated with the pattern of interest, but are not in fact causally related to it. Generating more precise predictions will require modeling efforts that are much more quantitative than most existing conceptual models in primatology. Acknowledging that most primatologists are not trained in mathematics, more precise predictions from existing conceptual models can be obtained readily by the use of individual- or agent-based models that are relatively accessible to those willing to learn the software (Bonnell *et al.* 2010; Sellers *et al.* 2007).

Time is important; the ecology of natural forests and their primates are changing rapidly through human-caused global and local changes (Estes *et al.* 2011; Janson 2000a). If we are to be able to infer the mechanisms that have shaped the emergence of and continued variation in behavior and ecology among primate species in relatively pristine conditions, we must use the strongest scientific methods available, as well as working to conserve the broadest sample of primate adaptive types. The recent advances in field and statistical methods described here, along with more quantitative modeling efforts, will allow us to maximize the scientific knowledge gained from our hard-won fieldwork.

**Acknowledgments** I am deeply grateful to the many assistants, students, and collaborators who have enabled the work described here over the past 39 yr. For the research in Iguazú National Park, Argentina, I make special note of the unfailingly excellent field support by Maria Celia Baldovino. Permission to conduct these studies was graciously provided by the Department of Forests and Wildlife of the Peruvian Ministry of Agriculture and by the Argentine National Park Service and the Delegacion Tecnica Regional Nordeste. Financial support for my research was provided by a series of NSF grants (BNS- 8007381, BNS-9009023, IBN-9511642, BNS-9870909, BCS-0515007), as well as a Fulbright fellowship and grants from Sigma Xi Grants-in-Aid-of-Research and National Geographic Society's Committee on Research and Exploration.

## References

- Altmann, J. (1974). Observational study of behavior: sampling methods. *Behaviour*, 49, 227–265.
- Bolker, B., Brooks, M., Clark, C., Geange, S., Poulsen, J., Stevens, M., et al. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135.

- Bollen, K. A., & Jackman, R. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In J. Fox & J. S. Long (Eds.), *Modern methods of data analysis* (pp. 257–291). Newbury Park: Sage.
- Bonnell, T. R., Sengupta, R. R., Chapman, C. A., & Goldberg, T. L. (2010). An agent-based model of red colobus resources and disease dynamics implicates key resource sites as hot spots of disease transmission. *Ecological Modeling*, *221*(20), 2491–2500.
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): the general theory and its analytical extensions. *Psychometrika*, *52*, 345–370.
- Cramer, A. E., & Gallistel, C. R. (1997). Vervet monkeys as travelling salesmen. *Nature*, *387*(6632), 464–464.
- Di Bitetti, M. S. (2001). *Food-associated calls in the tufted capuchin monkey (Cebus apella)*. Ph.D. dissertation, Stony Brook University, Stony Boork, NY. xvii + 258 p.
- Di Bitetti, M. S. (2003). Food-associated calls of tufted capuchin monkeys (*Cebus apella nigritus*) are functionally referential signals. *Behaviour*, *140*, 565–592.
- Di Bitetti, M. S. (2005). Food-associated calls and audience effects in tufted capuchin monkeys, *Cebus apella nigritus*. *Animal Behaviour*, *69*, 911–919.
- Di Bitetti, M. S., Vidal, E. M. L., Baldovino, M. C., & Benesovsky, V. (2000). Sleeping site preferences in tufted capuchin monkeys (*Cebus apella nigritus*). *American Journal of Primatology*, *50*(4), 257–274.
- Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., et al. (2011). Trophic downgrading of planet Earth. *Science*, *333*, 301–306.
- Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. W. (2009). Gestural communication of the gorilla (*Gorilla gorilla*): repertoire, intentionality and possible origins. *Animal Cognition*, *12*, 527–546.
- Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews* *41*: 587–640.
- Grether, G., Palombit, R. A., & Rodman, P. S. (1992). Gibbon foraging decisions and the marginal value model. *International Journal of Primatology*, *13*, 1–17.
- Hall, K. R. L., & DeVore, I. (1965). Baboon social behavior. In I. DeVore (Ed.), *Primate behavior: Field studies of monkeys and apes* (pp. 53–110). New York: Holt, Rinehart and Winston.
- Hamilton, W. D. (1971). Geometry for the selfish herd. *Journal of Theoretical Biology*, *31*, 295–311.
- Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology*, *4*(3), 194–205.
- Hausfater, G. (1975). *Dominance and reproduction in baboons (Papio cynocephalus)*. Basel: Karger.
- Heinrich, B., & Marzluff, J. M. (1991). Do common ravens yell because they want to attract others? *Behavioral Ecology and Sociobiology*, *28*, 13–21.
- Hewitt, J. E., Thrush, S. F., Dayton, P. K., & Bonsdorff, E. (2007). The effect of spatial and temporal heterogeneity on the design and analysis of empirical studies of scale-dependent systems. *American Naturalist*, *169*, 398–408.
- Hirsch, B. T. (2010). Tradeoff between travel speed and olfactory food detection in ring-tailed coatis (*Nasua nasua*). *Ethology*, *116*, 671–679.
- Janmaat, K. R. L., Byrne, R. W., & Zuberbuhler, K. (2006a). Evidence for a spatial memory of fruiting states of rainforest trees in wild mangabeys. *Animal Behaviour*, *72*, 797–807.
- Janmaat, K. R. L., Byrne, R. W., & Zuberbuhler, K. (2006b). Primates take weather into account when searching for fruits. *Current Biology*, *16*(12), 1232–1237.
- Janson, C. H. (1983). Adaptation of fruit morphology to dispersal agents in a Neotropical forest. *Science*, *219*(4581), 187–189.
- Janson, C. H. (1984). Female choice and mating system of the brown capuchin monkey *Cebus apella* (Primates, Cebidae). *Zeitschrift Tierpsychologie—Journal of Comparative Ethology*, *65*(3), 177–200.
- Janson, C. H. (1985). Aggressive competition and individual food consumption in wild brown capuchin monkeys (*Cebus apella*). *Behavioral Ecology and Sociobiology*, *18*(2), 125–138.
- Janson, C. H. (1986). The mating system as a determinant of social evolution in capuchin monkeys. In J. Else & P. C. Lee (Eds.), *Primate ecology and conservation* (pp. 169–179). Cambridge: Cambridge University Press.
- Janson, C. H. (1988a). Intra-specific food competition and primate social structure: a synthesis. *Behaviour*, *105*, 1–17.
- Janson, C. H. (1988b). Food competition in brown capuchin monkeys (*Cebus apella*): quantitative effects of group size and tree productivity. *Behaviour*, *105*, 53–76.
- Janson, C. H. (1990a). Ecological consequences of individual spatial choice in foraging groups of brown capuchin monkeys, *Cebus apella*. *Animal Behaviour*, *40*, 922–934.
- Janson, C. H. (1990b). Social correlates of individual spatial choice in foraging groups of brown capuchin monkeys, *Cebus apella*. *Animal Behaviour*, *40*, 910–921.

- Janson, C. H. (1996). Towards an experimental socioecology of primates: Examples from Argentine brown capuchin monkeys (*Cebus apella nigrinus*). In M. A. Norconk, A. Rosenberger, & P. Garber (Eds.), *Adaptive radiations of Neotropical primates* (pp. 309–325). New York: Plenum Press.
- Janson, C. H. (1998). Experimental evidence for spatial memory in foraging wild capuchin monkeys, *Cebus apella*. *Animal Behaviour*, *55*, 1229–1243.
- Janson, C. H. (2000a). Primate socio-ecology: the end of a golden age. *Evolutionary Anthropology*, *9*(2), 73–86.
- Janson, C. H. (2000b). Spatial movement strategies: theory, evidence, and challenges. In S. Boinski & P. Garber (Eds.), *On the move. How and why animals travel in groups* (pp. 165–203). Chicago: Chicago University Press.
- Janson, C. H. (2007a). Experimental evidence for route integration and strategic planning in wild capuchin monkeys. *Animal Cognition*, *10*(3), 341–356.
- Janson, C. H. (2007b). Predator detection and the evolution of primate sociality: insights from experiments on a rain forest primate. *American Journal of Physical Anthropology*, *S44*, 136–136.
- Janson, C. H., & DiBitetti, M. S. (1997). Experimental analysis of food detection in capuchin monkeys: effects of distance, travel speed, and resource size. *Behavioral Ecology and Sociobiology*, *41*(1), 17–24.
- Janson, C. H., & Goldsmith, M. (1995). Predicting group size in primates: foraging costs and predation risk. *Behavioral Ecology*, *6*, 326–336.
- Janson, C. H., & van Schaik, C. P. (1993). Ecological risk aversion in juvenile primates: Slow and steady wins the race. In M. Pereira & L. Fairbanks (Eds.), *Juvenile primates: Life history, development and behavior* (pp. 57–74). New York: Oxford University Press.
- Janson, C. H., & Vogel, E. R. (2006). Hunger and aggression in capuchin monkeys. In G. Hohmann, C. Boesch, & C. Nebel (Eds.), *Feeding ecology in apes and other primates* (pp. 285–312). Cambridge: Cambridge University Press.
- Kappeler, P., & Watts, D. (Eds.). (2012). *Long-term studies of primates*. Heidelberg: Springer.
- Mundry, R., & Nunn, C. L. (2009). Stepwise model fitting and statistical inference: turning noise into signal pollution. *American Naturalist*, *173*, 119–123.
- Pickett, S. T. A., Kolasa, J., & Jones, C. G. (2007). *Ecological understanding: The nature of theory and the theory of nature*. Amsterdam: Elsevier Press.
- Platt, J. R. (1964). Strong inference. *Science*, *146*(3642), 347–353.
- Sagarin, R., & Pauchard, A. (2010). Observational approaches in ecology open new ground in a changing world. *Frontiers in Ecology and the Environment*, *8*, 379–386.
- Sellers, W. I., Hill, R. A., & Logan, B. S. (2007). An agent-based model of group decision making in baboons. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362* (1485), 1699–1710.
- Sokal, R. R., & Rohlf, F. J. (1995). *Biometry: The principles and practice of statistics in biological research*. New York: W. H. Freeman.
- Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton: Princeton University Press.
- Terborgh, J. (1983). *Five New World primates: A study in comparative ecology*. Princeton: Princeton University Press.
- Vine, I. (1971). Risk of visual detection and pursuit by a predator and the selective advantage of flocking behaviour. *Journal of Theoretical Biology*, *30*, 406–422.
- Visalberghi, E., Janson, C. H., & Agostini, I. (2003). Response toward novel foods and novel objects in wild *Cebus apella*. *International Journal of Primatology*, *24*(3), 653–675.
- Wheeler, B. C. (2008). Selfish or altruistic? An analysis of alarm call function in wild capuchin monkeys, *Cebus apella nigrinus*. *Animal Behaviour*, *76*, 1465–1475.
- Wheeler, B. C. (2009). Monkeys crying wolf? Tufted capuchin monkeys use anti-predator calls to usurp resources from conspecifics. *Proceedings of the Royal Society of London B: Biological Sciences*, *276* (1669), 3013–3018.
- Wheeler, B. C. (2010a). Decrease in alarm call response among tufted capuchins in competitive feeding contexts: possible evidence for counterdeception. *International Journal of Primatology*, *31*(4), 665–675.
- Wheeler, B. C. (2010b). Production and perception of situationally variable alarm calls in wild tufted capuchin monkeys (*Cebus apella nigrinus*). *Behavioral Ecology and Sociobiology*, *64*(6), 989–1000.
- Zuberbuhler, K., Jenny, D., & Bshary, R. (1999). The predator deterrence function of primate alarm calls. *Ethology*, *105*(6), 477–490.