EFFECTS OF AGGRESSIVE INTERACTIONS ON ANTIPREDATOR BEHAVIOR: EMPIRICAL AND THEORETICAL ASPECTS

By

Ramón Díaz-Uriarte

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Zoology)

at the

UNIVERSITY OF WISCONSIN – MADISON

2000

Contents

Abstract

Avoiding predators may conflict with territorial defense because a hiding territorial resident is unable to monitor its territory or defend it from conspecific intrusions. With persistent intruders, the presence of an intruder in the near past can indicate an increased probability of future intrusions. Therefore, following a conspecific intrusion, territorial residents should minimize costs from future intrusions at the cost of higher predation risks. The main focus of this thesis is to investigate changes in antipredator behavior following a conspecific intrusion.

In the first chapter I examine the existence of effects of past conspecific intrusions on antipredator behavior and how these effects differ from the changes in antipredator behavior related to the immediate (vs. past) presence of a conspecific intruder. I conducted experiments with males of the territorial lizard, *Tropidurus hispidus*, recording approach distance (distance between predator and prey when prey escapes) and time to re-emerge from a refuge after hiding. Past aggressive interactions affected antipredator behavior: lizards re-emerged sooner (compared to a control) when the predator attacked 5 min after an aggressive encounter. If the predator attacked while an aggressive encounter was ongoing, there was also a reduction in approach distance. The results: (1) are consistent with an economic hypothesis that predicts that *T. hispidus* incur greater predation risks to minimize future territorial intrusions; (2) show that effects of past and ongoing aggressive interactions are different, consistent with minimization of present intrusion costs.

In the second chapter I investigate whether testosterone manipulations affect antipredator behavior and the effects of past aggressive interactions. Elevated testosterone levels in lizards result in males that increase their allocation to territorial defense at the expense of other costs. Consequently, we expected that elevated testosterone would: (1) increase exposure to predation; (2) produce a disproportionate increase in exposure to predation following a past aggressive interaction. We manipulated testosterone levels of male *T. hispidus* using subcutaneous testosterone implants. Our results provide strong evidence that past aggressive interactions result

in increased exposure to predation and that the type of first encounter (aggressive interaction with a conspecific vs. control presentation) had long-lasting effects on antipredator behavior. We found no evidence of differences in aggressive behavior related to hormonal treatment, of an association between aggressive and antipredator behaviors, or of an increase in exposure to predation with increased testosterone level. The lack of effects of testosterone on antipredator behavior could be the consequence of testosterone manipulations not altering aggressive behavior on males of this species, a pattern that might not be uncommon in tropical vertebrates.

In the third chapter I use a mathematical model to examine the effects that past conspecific intrusions can have on antipredator behavior, when intruders are persistent, focusing mainly on the effects of rate of intrusion of other conspecifics, the behavior of the reintruder, and the timing of the predator's attack. Past aggressive intrusions rarely affect time to hide: the optimal behavior is to hide as soon as the predator initiates its attack. Time to reemerge is strongly affected by past aggressive interactions (animals reemerge sooner from a refuge), and these effects depend on the time of the predator's attack, the reintruder's pattern of return, and the intrusion rates of other conspecifics. Differences between my findings and those from previous studies suggest that the trade-off between antipredator behavior and territorial defense can involve different types of costs than the trade-off antipredator behavior-foraging.

Together, these chapters are relevant for studies of the changes in antipredator behavior due to changes in the social environment, and they establish a connection between population level processes, mating system and defensibility of resources, and antipredator behavior. These three chapters can have empirical and theoretical relevance for studies of the costs, (co)evolution, and ecological consequences of territorial and antipredator strategies.

In the first two chapters I use cross-over designs extensively. These types of designs are frequently used in animal behavior studies as they allow experiments with relatively small numbers of subjects that nonetheless achieve high statistical power by using each subject as its own control. However, cross-over trials are often analyzed incorrectly in the behavioral

literature, and many statistics textbooks used by behaviorist either do not mention them or contain potentially misleading advice. Moreover, some of my experiments involve data, such as multivariate responses and censored observations, which although common in many behavioral experiments are not generally considered in detail in statistical textbooks on cross-over trials. The last two chapters address these issues.

In chapter four I review the use of cross-over trials in the behavioral literature, and I explain why the traditional analyses (based on paired t-tests) are inappropriate, the problems associated with carry-over effects, and the types of cross-over designs that are potentially most useful for behaviorists. In the fifth chapter I review methods of analyses of cross-over trials in the context of animal behavior experiments. I group methods of analysis according to the type of response variable: non-parametric and robust methods for metric responses, parametric methods for metric responses —linear mixed-effects models—, models for categorical responses both non-parametric and parametric —extensions of generalized linear models—, censored observations —survival analysis–, and multivariate responses. Within-individual contrasts are explained in detail as they are the basis of many different methods, from non-parametric to multivariate and survival-based models, and they offer a useful framework for extending the analysis of data from cross-over trials to situation where robust methods might be needed (e.g., permutation tests of censored multivariate responses). In this chapter I also discuss some types of plot that are specific and particularly useful for cross-over trials. If design, wash-out periods, and type of response are given the appropriate consideration, cross-over designs can be very powerful tools for behaviorists whenever obtaining new subjects is more costly than repeatedly testing the same individual, and thus in particular for researchers working in the lab or in field enclosures where animals require lengthy training or habituation.

Acknowledgments

Each chapter contains an acknowledgments section, which attempts to give credit to the people (and institutions) whose help I enjoyed during each particular project. Here, I would like take the opportunity to thank those whose help goes beyond any single particular chapter.

During the first two years of grad school I was supported by a grant from the La Caixa fellowship program. A John and Virginia Emlen Graduate Student Fellowship from the Department of Zoology, University of Wisconsin-Madison, provided me with salary for a summer and covered the costs of publication of the first chapter of this thesis. I have also benefited (for three times) from the John Jefferson Davis travel award (Dept. Zoology, University of Wisconsin-Madison) to cover expenses of attendance to meetings.

I would like to thank my advisor, Cathy A. Marler, and the members of my PhD committee, David H. Abbott, Jeffrey R. Baylis, Anthony R. Ives, Erick V. Nordheim, Charles T. Snowdon, for advice and encouragement during all the stages of this work. In particular, Chuck provided advice while he visited us in Brazil and discussed with me the use of enclosures for my field work; he has also always shown interest (or was it just amusement?) in my flying lizards engineering projects. Thanks to Tony's advice I have become much more critical of my own theoretical work, and about when models are, or are not, worth being done. I owe Rick pursuing degrees in statistics, and I am thankful for his patience and his advice during some difficult periods early on in my graduate career. Cathy accepted the challenge of having me as her first grad student, and has been very supportive during these years, never complaining about my drifting away from endocrinology. Moreover, she has always encouraged me to pursue my own interests (such as my MS in statistics), not matter how little that overlapped with what everybody else was doing in the lab; I thank her for that.

All of the empirical work in this thesis was carried out in Brazil. I thank the Departamento de Fisiologia, Universidade Federal do Rio Grande do Norte, Brazil, for logistic support and use of facilities, and in particular M. Fátima Arruda, M. Bernardete C. de Sousa, and Arrilton

Araujo for help in a myriad of ways, from finding a place to live, to dealing with customs, or arranging help of kids for lizard capture. Without their help my empirical work would not have been possible, and our life in Brazil would have been a lot more complicated. I was very fortunate to count with their assistance. Almost all of the work reported here was conducted at the Eflex-IBAMA of Nisia Floresta, and I am thankful for permission to work at this site. In particular, I would like to acknowledge the help of João Dantas for his patience with my enclosures and understanding of the painted lizards. Edinaldo de Nascimento taught me some nice tricks about noosing *Tropidurus*; in addition, he provided help (and friendship) during the study, and helped me get in touch with several kids who collaborated in lizard capture. I would also like to thank the owners of the Fazenda Luis Carlos (Santo Antonio, RN) for permission to work there.

While in Brazil we had the privilege of enjoying the friendship of Maria Carla Lopes de Nascimento, Rebeca Harrison, Fabiola Albuquerque and Carla S. de Castro; yes, we miss them. M. Elisa Freire, Adalberto Varela, Dwain Santee, and specially Antonio C. de Andrade Moura generously showed me a variety of places in the northeast of Brazil to go herping. I would also like to acknowledge the hospitality and friendliness that I received while visiting Sitio Bravo (Paraiba), the Biological Station of the UFPB in S. J. de Carirí (Paraiba), and the Ecological Station of Seridó. Before leaving for Brazil I received very valuable information and reprints about *Tropidurus* and other lizards from Michaela Heisig, Carlos F. Duarte da Rocha, Miguel T. Rodrigues, Monique VanSluys, Laurie J. Vitt, and Pete Zani; I thank them for helping a complete stranger that bugged them with questions about *Tropidurus*, how abundant they were, how easy they were to catch, whether they were better for my work than *Anolis*, etc., etc. Brazil, in general, was quite an experience in many different ways. Everyday it struck me that I was using a piece of equipment (e.g., an HP calculator) whose price was about that of the monthly minimum wage in Brazil. This has given me food for thought. But I have the feeling that I got from Brazil a lot more than I gave in return.

I took my first steps in science with Dirk Bauwens (GJSSSI), who taught me things about science I have not forgotten, and who has remained a friend and a source of wise advice over the years. I hope that now I'll qualify to be introduced to some of his illustrious acquaintances, responsible for Dirk's love of simple proportionality. Elliott Sober has had a deep influence on the way I see science and evolutionary biology. Thomas Kurtz taught a class on stochastic processes that fascinated me and tried to convince me that I had to add some math rigor to my intuition. I have worked with Ted Garland on the comparative method and I have also learned from him about giving talks and writing papers. Steve Adolph generously accepted to show me —a stranger— the lizards and questions he was working on, and over the years has provided very helpful advice on matters of science and grad school. Douglas Bates introduced me to the GNU/Linux operating system and to open source, free software —free like in $0, but specially free like in freedom and available to all; R. M. Stallman makes this point much better than I do. This has changed the way I use and think about software, and made me a convert (leaving behind "the other operating system" for good) . In particular, I have benefitted immensely from R, a statistical computing system "not unlike S", and LyX, a LaTeXbased document processor, as well as from their helpful communities of users.

While in Madison, I have enjoyed the friendship of Janet Bester, Susie Bulova, Mark Coleman, Kevin Gross, Charlie Meredith, John Swallow, Mariano Sironi, Stella de la Torre, and Brian Trainor. They have made life in a place without mountains a lot more livable. Which brings me to Doug Hemken and specially Hobby Martini for being great rock climbing partners. Also while in Madison I discovered Thich Nhat Hanh's work.

From my family (in particular my parents, brothers, and aunt) I have received much support (moral, economic, logistic, bureocratic, and otherwise) from the very beginning, even when they probably thought I could be doing better things with my time. I have underestimated this support in the past; and I might still be underestimating it, since doing my thesis meant being here while they were there; but they kept it quiet, asking instead questions about the work

itself (and I know they have read some of my papers). With their actions, they have made it clear that they cared, and I am little bit embarrased to recognize that they are more likely to remeber when my defense is than I am to remember when any of their birthdays are. It feels good to think about going back.

Cristina Lázaro-Perea shows up in every one of the acknowledgments because she has been deeply involved in this thesis in many ways. Over the years, discussions with her have strongly influenced my work; in fact, the main topic of this dissertation (how aggression affects antipredator behavior) grew out of a discussion where she kept insisting that what I regarded as just a neat side observation ("Los *Tropis* siempre me dejan acercarme más cuando se están pegando; estoy seguro de que este efecto dura un rato largo") was really a lot more interesting (and feasible) than all that loser/winner stuff I was trying to work on. Cristina has also provided invaluable material and logistic help in the field (building and cleaning enclosures, for example), and countless revisions of manuscripts (which has improved my writing). In Brazil she was an example of perseverance and commitment that set a very high standard for me. Cristina's help extends beyond the professional, although after 10 years together there could be a risk of taking many things for granted. So I won't: I want to thank her for her generosity, support, patience, and honesty that have helped keep me going during these seven years. In the meantime, she has made a better person out of me. I am immensely fortunate to share my life with her.

I was not fully aware what I was giving up when I started grad school. From being there while my younger brother went from childhood to young adulthood or my father retired, to seeing my newborn niece, or the long summer nights with friends, or the honest feeling of usefulness of political activism, or the mountains, the sun and the skies of the landscapes I love. I dedicate this thesis (ex post facto) to all those whom I lost a little bit along the way, and to the dreams and desires that got postponed.

List of publications

(Note: this list is updated only up to early 2001, and includes only already published papers or books).

Díaz-Uriarte, R. 2001. Territorial intrusion risk and antipredator behaviour: a mathematical model. Proceedings of the Royal Society of London, Series B, 268, 1165-1173.

Díaz-Uriarte, R. 1999. Anti-predator behaviour changes following an aggressive encounter in the lizard Tropidurus hispidus. Proceedings of the Royal Society of London, Series B, 266: 2457-2464.

Garland, T. Jr., & Díaz-Uriarte, R. 1999. Polytomies and independent contrasts: an examination of the bounded degrees of freedom approach. Systematic Biology, 48: 547-558.

Díaz-Uriarte, R. 1998. Increased risk of predation following an aggressive encounter. a mathematical model. American Zoologist (abstract), 38: 168A.

Díaz-Uriarte, R., & Garland, T. Jr. 1998. Effects of branch length errors on the performance of phylogenetically independent contrasts. Systematic Biology, 47: 654-672.

Garland, T. Jr., Martin, K. L. M., & Díaz-Uriarte, R. 1997. Reconstructing ancestral trait values using squared-change parsimony: plasma osmolarity at the origin of amniotes. Pp. 425-501 in: S. S. Sumida & K. L. M. Martin (eds.), Amniote origins: completing the transition to land, Academic Press, San Diego.

Bauwens, D., & Díaz-Uriarte, R. 1997. Covariation of reproductive traits of lacertid lizards: a comparative study. The American Naturalist, 149: 91-111.

Díaz-Uriarte, R, & Garland T. Jr. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. Systematic Biology, 45: 27-47.

Díaz, J. A., Díaz-uriarte, R., Rodríguez, A. 1996. Influence of behavioral thermoregulation on the use of vertical surfaces by Iberian wall lizards, Podarcis hispanica. Journal of Herpetology, 30: 548-552.

Díaz-Uriarte, R, Bauwens, D, & Barbadillo, L. J. 1994. Covariation of reproductive traits in lacertid lizards: a comparative study. American Zoologist (abstract), 34: 6A.

Díaz-Uriarte, R, & Díez, I. 1985. Las Aves Urbanas. Penthalon. Madrid.

Chapter 1

Antipredator behaviour changes following an aggressive encounter in the lizard *Tropidurus hispidus*

This chapter has been published in the
Proceedings of the Royal Society of London, Series B, 1999, 266, 2457-2464

1.1 Abstract

Avoiding predators may conflict with territorial defence because a hiding territorial resident is unable to monitor its territory or defend it from conspecific intrusions. With persistent intruders, the presence of an intruder in the near past can indicate an increased probability of future intrusions. Therefore, following a conspecific intrusion, territorial residents should minimise costs from future intrusions at the cost of higher predation risks. I conducted experiments with males of the territorial lizard, *Tropidurus hispidus*, recording approach distance (distance between predator and prey when prey escapes) and time to re-emerge from a refuge after hiding. Past aggressive interactions affected antipredator behaviour: lizards re-emerged sooner (compared to a control) when the predator attacked 5 min after an aggressive encounter. If the predator attacked while an aggressive encounter was ongoing, there was also a reduction in approach distance. The results: a) are consistent with an economic hypothesis that pre-

dicts that *T. hispidus* incur greater predation risks to minimise future territorial intrusions; b) show that effects of past and ongoing aggressive interactions are different, consistent with minimisation of present intrusion costs. These results are relevant for studies of the changes in aggressive behaviour due to changes in the social environment, and for studies of the costs and (co)evolution of aggressive and antipredator strategies.

1.2 Introduction

Optimal antipredator behaviour should be the result of weighting the risk of predation against the benefits from other activities. Experimental and theoretical work, focused mainly on the trade-off between foraging and predator avoidance, has shown that changes in the terms of the trade-off between mortality risk from predation and costs of hiding/escaping from predators will change the behavioural optimum (see Clark 1994; Ydenberg & Dill 1986; reviews in Lima & Dill 1990; Lima 1998). Thus, when the costs of interrupting other activities increase (e.g., foraging at a better patch or consuming a larger prey), animals adopt behavioural strategies that lead to increases in risk of mortality from predation (e.g., delaying escape from a predator or reemerging sooner from a refuge). In territorial animals, territorial defence can be an important determinant of reproductive success. However, compared to the antipredator-foraging trade-off, there is little information about trade-offs between antipredator behaviour and territorial defence. The general aim of this study was to examine how predation-related risk taking behaviour changes as a consequence of past and present aggressive interactions that increase territorial costs of hiding; the two hypotheses tested predict increased exposure to predation as a consequence of increased costs of hiding due to past (first hypothesis) or present (second hypothesis) territorial conspecific intrusions.

A predatory attack creates conflicting demands on a territorial animal: hiding decreases risk of mortality from predation, but minimises the chances of detecting and repelling a conspecific intruder (i.e., increases the territorial costs of hiding). These territorial costs of hiding can

be specially high following a conspecific intrusion: in some territorial species intruders obtain or enlarge territories by persistently intruding into the territories of settled animals (review in Stamps & Krishnan 1995, 1998; e.g., lizard *Anolis aeneus*: Stamps & Krishnan 1995; red-winged blackbirds: Yasukawa 1979; purple martins: Stutchbury 1991; song sparrows: Arcese 1987). Thus, the occurrence of one aggressive encounter can inform a territorial resident that subsequent territorial intrusions are likely.

The first hypothesis tested in this study states that a past territorial intrusion changes the terms of the trade-off between predation and vigilance by increasing the territorial costs of hiding, and thus alters the behavioural optimum. Therefore, if a predator attacks soon after an aggressive interaction is over, a territorial resident should modify its behaviour to decrease the chances of territorial intrusions at the cost of increased predation risks (hereafter called extended effects of aggression on antipredator behaviour). The predictions from this hypothesis are that, following an aggressive encounter, a territorial resident will show a decrease in the distance at which it flees from a predator and/or a decrease in the time till it re-emerges from a refuge after the predator attacks. These predictions were tested in Experiment 1 using a human as a simulated predator and comparing antipredator behaviour in males of the lizard *Tropidurus hispidus* 5 min after the end of an aggressive interaction to antipredator behaviour 5 min after a control presentation.

The antipredator behaviour consequences of a change in the territorial costs of hiding can be further studied by examining the difference between the effects of an aggressive encounter that has finished (extended effects) and an ongoing aggressive interaction (immediate effects). In an ongoing aggressive encounter the intruder is in the territory when the predator attacks, and hiding could result in much larger intrusion costs, specially if the approaching predator is not an attacking one. The second hypothesis tested in this paper states that current presence of an intruder increases territorial costs of hiding with respect to past presence of an intruder, and thus territorial residents should show further increases in their exposure to predation when the

predator approaches during an ongoing aggressive encounter vs. sometime after the end of the aggressive interaction. This hypothesis predicts that immediate effects will result in a decrease in the distance at which the territorial resident flees from a predator and/or a decrease in the time till it re-emerges from a refuge after the predator attacks compared to extended effects. I examined this hypothesis in Experiment 2 by comparing antipredator behaviour of male *T. hispidus* during an ongoing aggressive encounter to antipredator behaviour 5 min after the end of the aggressive interaction.

1.3 Methods

1.3.1 Animals and study site

Experiments (Table 1.1, p. 8) were conducted at the Nisia Floresta Forest Experimental Station, EFLEX-IBAMA, (6° 5' S, 35° 12' W), located 45 Km from Natal (Northeastern Brazil); Experiment 1 (hereafter Exp. 1) was conducted between 27-April and 22-May, 1997, and Experiment 2 (hereafter Exp. 2) between 29-November-1997 and 13-January-1998. I used adult males of the lizard *Tropidurus hispidus* (snout-vent length [SVL] 70-130 mm), a widespread, diurnal, sit-and-wait iguanine lizard in South America (Rodrigues 1987; Vitt 1995). In the area studied both male and females were territorial through the year, and encounters among males that developed into escalated fights tended to repeat themselves (with the same contenders) in subsequent hours/days (pers. obs.).

Experimental subjects were adult males (SVL$\geq$ 100 mm), captured in villages close to the station, that had not been used in other experiments, or used before as intruders, or later used as intruders in the same enclosure. Intruders (adult males SVL $>$ 90 mm) were used a maximum of three times and were never wounded by the experimental procedure. The same experimental animal was not exposed to the same intruder more than once. Intruders were assigned at random to experimental animals, but no intruder could be used twice in the same

enclosure and for the same treatment (in Exp. 2). Moreover, for each experimental animal in Exp. 2, none of the two treatments could be applied using either the two largest or the two smallest intruders, to ensure adequate interspersion with respect to intruders' sizes (this is not applicable to Exp. 1 where each experimental animal was subject to only one intruder). All animals were released in the area of capture at the end of testing.

1.3.2 Enclosures and animal husbandry

I used enclosures to minimise variation in behaviour. Enclosures were located in open patches in plantation areas and measured 3.6 to 4.9 m^2 (2 to 2.5 by 2 m) in Exp. 1 and 4 m^2 (2 by 2 m) in Exp. 2. Enclosures were 1 m high, constructed from transparent plastic, sunk 15 cm into the ground, attached to a wood frame. Each enclosure contained two refuges made with bricks and roof tiles that offered protection and were readily used by the lizards as hiding places. Enclosures were partially covered from above to provide shade during the central hours of the day. Enclosures also included one or two females (and in some cases one small male; see Table 1.1 (p. 8)). All females were randomly assigned to enclosures/males, except that females' SVL had to be at least 5 mm less than the males' (in the field, males were associated with smaller females).

I placed a blind 7.5 m away from the enclosure. Using suspended fishing lines, I could move an intruder from behind the blind to inside the enclosure and retrieve it at the end of the trial, without my ever leaving the blind. When I approached the enclosures for feeding or small repairs I used a poncho which contrasted with the clothes used during tests (white pants and T-shirt).

Enclosures were more than 15 meters apart with dense and tall intervening vegetation ensuring no visual contact between them, and were placed in areas where, during a period of ten months, I only observed four free-ranging adult *T. hispidus* (one male, three females). Thus, interactions with naturally-occurring conspecifics should have been extremely rare.

Lizards were fed every two to three days a diet of crickets, mealworms, fly maggots, roaches and beetles, and a mixture of egg, powdered milk, and fruit. In Exp. 1 water was available naturally (rainy season) and animals were fed one or two days before testing started, and were not fed during the days of testing. In Exp. 2 (dry season), enclosures had several water containers, and animals were fed one or two days preceding testing, and early on the third day or, after testing, on the second day. Enclosures were cleaned of faecal boli before introducing new experimental animals.

Animals in the enclosures displayed normal antipredator behaviour: *T. hispidus* uses refuges for hiding when a predator attacks (Vitt 1995) and in the study area I observed wild *T. hispidus* run into refuges when attacked by the predators dogs, cats, chickens, and common marmosets (*Callithrix jacchus*), and when potential predators (e.g., crane hawk, *Geranospiza caerulescens*, caracara, *Polyborus plancus*) flew over. Moreover, in this region of Brazil, *T. hispidus* are very frequently killed by humans (particularly children). *T. hispidus* in the enclosures not only sought refuge when approached by a human, but also when crane hawks and caracaras flew over.

Animals in the enclosures also displayed normal aggressive and mating behaviour: males attacked intruders, and courted and mated with females; more than nine females laid eggs and at least six clutches hatched successfully in the enclosures. Body mass did not change between the time the animals were introduced and the time they were removed from the enclosures (Exp. 1: mean change [final-initial mass] $\pm$ s.e.$= -0.27 \pm 0.409$ g, paired $t_{14} = 0.67$, $p = 0.512$; Exp. 2: mean change $\pm$ s.e.$= 1.33 \pm 0.736$ g, paired $t_{11} = 1.89$, $p = 0.085$). While in the enclosures, lizards were rarely approached by humans (except myself).

1.3.3 Experimental design and antipredator tests

In both experiments, animals were tested several days (Table 1.1, p. 8) after being introduced in an enclosure to ensure that animals were used to the enclosures. I used cross-over designs

(Jones & Kenward 1989): each animal was subject to two treatments through time, so that treatment differences are estimated using within-animal comparisons. Each animal received only one treatment per day, in the sequences shown in Table 1.1 (p. 8), and was tested in successive days and at about the same hour as the preceding day. Thus, the testing phase lasted two days for each animal in Exp. 1, and four days for each animal in Exp. 2. Both experiments involved presenting a male lizard with a stimulus (intruder or control) and, some time later, measuring antipredator behaviour by simulating a predatory attack. A test (stimulus presentation + antipredator test) lasted approximately 40 min per animal.

In Exp. 1 I measured antipredator behaviour 5 min after an intruder encounter (E: extended effects) and 5 min after a control (C) presentation. In Exp. 2 I measured antipredator behaviour during an ongoing aggressive interaction with an intruder (I: immediate effects) and 5 min after the end of the interaction (E: extended effects). Details of the experiments are shown in Table 1.1 (p. 8). When escaping predators *T. hispidus* need to decide when to flee from the predator and, after hiding, when to re-emerge from the refuge; thus, the variables measured were chosen to reflect these two decissions and are explained in Table 1.2 (p. 9). To run the antipredator test, I positioned myself 13 m away from the enclosure (4.5 m behind the blind) and approached the lizard directly at a moderate speed (Exp. 1: mean = 0.42 m/s, s.d. = 0.056; Exp. 2: mean = 0.46 m/s, s.d. = 0.047). Whenever the lizard moved, I stopped for 15 sec and recorded my position, and then approached again. The approach-and-stop continued until the lizard hid, when I moved to a spot at a fixed distance from the enclosure (Exp. 1, 2 m; Exp. 2, 4.5 m), and remained motionless for 20 min. I recorded my movements and the lizard's behaviours using an HP-48GX calculator for continuous event recording. All tests were conducted when lizards were active and air temperature (shaded bulb at 1.5 m) was higher than 26 °C.

Animals were habituated to the movement of the intruder delivery system using a toothpaste container (to prevent habituation to the control) with which I mimicked the movements I would use during the intruder and control presentations. Lizards were subject to 4 to 10 habituation

Table 1.1: Experiments 1 and 2: methods.

Exp.	Treatments	Sequences [1]	Subjects
1	Extended (E): - Introduced intruder male. - Left in enclosure max. 15 min. - Once attacked, left for 3 min and until three attacks. - Remove intruder. - Antipredator test; time end of intruder presentation to antipredator test: 5 min. Control (C): - Introduced wood stick ($\simeq$colour and size of adult male). - Left in enclosure for 3 min 45 sec[2]. - Remove control. - Antipredator test; time end of control presentation to antipredator test: 5 min.	EC,CE	- Three batches of six enclosures each. One experimental male per enclosure. - Each enclosure also two females (four enclosures) or one female and one small male (two enclosures)[3]. Females and small males the same in each enclosure throughout the experiment. - Experimental males assigned randomly to enclosures. - Three males in each batch assigned randomly to each sequence. - Males tested after 6 to 7 days in enclosures. - Sample size: 15 males[4].
2	Extended (E): - Introduced intruder male. - Left in enclosure max. 15 min. - Once attacked, left for 2 min (and a minimum of four attacks) or until six attacks, whichever came first. - Remove intruder. - Antipredator test; time end of intruder presentation to antipredator test: 5 min. Immediate (I): - Introduced intruder male. - Left in enclosure max. 15 min. - Once attacked, left for 2 min (and a minimum of four attacks) or until six attacks, whichever came first. - Antipredator test; i.e., intruder still within enclosure. - Intruder removed immediately after lizard hid[5].	EIIE,IEEI	- Six different enclosures used repeatedly, no batches . - One female and one experimental male introduced simultaneously in each enclosure (i.e., different females for each male). - Males assigned randomly to enclosures. - First animal tested assigned sequence at random; successive animals assigned immediately (before testing) alternating sequences. - Males tested when habituated (after 5 to 12 days in enclosures). - Sample size: 12 males[6].

[1]A sequence is the order in which the within-individual treatments are applied. An animal is assigned to a sequence, and treatments applied in the specified order (e.g., for sequence EC in Experiment 1 first testing day is E, second testing day is C). Therefore, Experiment 1 consisted of 2 periods and Experiment 2 of 4 periods, where a period is each one of the testing days.

[2]Median time that an intruder spent in enclosure in preliminary trials.

[3]In the field, a male's territory overlaps the territory of one or more females and often the home range of one or more small males. I never observed aggressive interactions between the experimental male and the small male.

[4]One of the enclosures could only be used during the first week and one animal was excluded from the study because it was hiding continuously during the day of testing.

[5]I obtained data for all four periods for all animals except two, one from each of the sequences.

[6]In the I-treatment removing the intruder from the enclosure took 1 min and involved some movement of the intruder-delivery-system. To control for these effects, in the E-treatment after the animal hid I approached the enclosure and remained next to it for 1 min, while moving the intruder delivery system to mimic the effects of removing an intruder.

Table 1.2: Response variables used to measure antipredator behaviour[1].

Variable	Description
Approach Distance	Distance between observer and the lizard when the lizard first initiated flight.
Minimum Distance	Minimum distance between the observer and the lizard before it initiated flight; the same as Approach Distance if there is only one run.
Time to Reemerge	Time since the lizard hid until it re-emerged (at least all the head was visible out of the refuge).
Time to Full Exposure	Time since the lizard hid until it was fully exposed (all the lateral surface of the body –not including tail– was visible out the refuge). Lizards in full exposure were generally more than one body length away from the refuge, they were visible (from many sight points) to both other lizards and potential predators, and were able to monitor their whole territory.

[1] The predictions tested refer to increases in predation risk that result from behavioural changes of the prey. As I could not measure predation risk directly I used the four response variables as proxies (and assumed that the risk of being killed is a decreasing function of each of the response variables). Approach Distance and Minimum Distance are proxies for risk when predator attacks; Time to Reemerge and Time to Full Exposure are proxies for risk at re-emergence. Thus, the four variables belong to two groups: initial attack and reemergence; results within each pair of variables should be consistent (i.e., either none of the two variables will depart from the null hypothesis, or the two variables will depart from it in the same direction).

trials, and were considered habituated if they did not hide during two successive habituation trials. In Exp. 2 I initially habituated some animals by hanging soda bottles for 24 to 48 h next to the enclosures (using the intruder delivery system); later, these animals were checked for habituation using the toothpaste container.

1.3.4 Statistical analyses

In Exp. 1 I analysed Approach Distance and Minimum Distance (Table 1.2 (p. 9)) with linear mixed-effects models, using the parameterisation in Jones & Kenward (1989, p. 30), but also

including several covariates and random effects. The full model examined was:

$$y_{ijklm} = \mu + \lambda_i + \beta X_j + \alpha_k + c_{j|k,\beta} + w_l + s_{jl} + \pi_m + \tau_{n[i,m]} + (\tau\beta)_n X_j + (\tau\alpha)_{kn} + (\alpha\beta)_k X_j + (\tau\alpha\beta)_{kn} X_j + e_{ijlm}, \quad (1.1)$$

where in the fixed effects part μ is the intercept, λ is the carry-over (which in this parameterisation is equivalent to a sequence effect), β is the coefficient for enclosure area (X), α is type of enclosure (two females or one female and one small male), π is the period effect (a period is each one of the occasions on which a treatment is applied, for example first or second day), τ is the direct treatment effect, and the terms in parentheses are interactions. In the random effects part c, w, and s are the random effects of enclosure, week, and individual respectively, and e are the within individual errors. All random effects are normal and independent of each other. When analysing Approach Distance I included my approach speed and the interaction approach speed* treatment. For the univariate analyses of Exp. 2 (all four variables –Table 1.2, p. 9) I used the linear mixed model

$$y_{ijkm} = \mu + \xi_i + c_j + (\xi c)_{ij} + s_{ijk} + \pi_m + \tau_{n[i,m]} + \lambda_{n[i,m-1]} + e_{ijkm} \quad (1.2)$$

where all terms are as in the model for Exp. 1, except for ξ which denotes sequence (sequence is the order in which the within-individual treatments are applied). Model fitting proceeded as in Exp. 1, except: a) I modelled the variance-covariance matrix of the within-individual errors e (examining the fit of compound-symmetric, autoregressive, general –unstructured positive definite–, and heteroscedastic error structures), because the data are repeated (>2) measures of the same individual; b) if period (as categorical variable) was left in the model, I attempted to simplify this structure by fitting linear and quadratic terms of period as a continuous variable. To fit these models I proceeded as explained in Pinheiro & Bates (2000), Diggle et al. (1994), and Littell et al. (1996).

In Exp. 1, for Time to Reemerge and Time to Full Exposure, nine and five, respectively, out of 30 (i.e., about 1/6 and 1/3) of the observations were right-censored (i.e., at 20 min the lizards still had not re-emerged or fully re-emerged), and thus require the use of techniques for censored data. I used the (first) approach suggested in Feingold & Gillespie (1996) after log-ranking (e.g., Lawless 1982, p. 420) the observations. To obtain p-values I used systematic permutation tests (Edgington 1995). In Exp. 2 Time to Reemerge and Time to Full Exposure had only a few right censored observations (two and seven, respectively, out of 46). Although residual plots did not indicate any problem with the models, I also analysed these data with the method of Feingold & Gillespie (1996), analogous to Exp. 1.

In both experiments I measured four response variables (Table 1.2, p. 9). To prevent inferential errors from four univariate tests of potentially correlated response variables, and to test for overall differences in antipredator behaviour taking into account the covariation among response variables, I used the multivariate permutation test for cross-over designs of Johnson & Mercante (1996). To give equal weights to all variables I scaled them to a mean of zero and variance of one before computing within-subject contrasts. (Simulations [Díaz-Uriarte & Nordheim, in prep.] indicate that the Type I error rate of the multivariate test with log-ranked censored data is the nominal one). I obtained the p-value for this test using systematic data permutation.

Permutation and multivariate tests were performed with code written in SPlus v. 3.3 (Statistical Sciences 1995). For Exp. 1, in all permutation tests animals were reassigned to sequences only within batch; for weeks two and three the permutation was conditional on the pattern of missing data. Mixed models were fitted using the SPlus library nlme (Pinheiro & Bates 2000) and SAS's PROC MIXED (Littell et al. 1996). All p-values are two-sided.

1.4 Results

1.4.1 Experiment 1: extended effects of aggression on antipredator behaviour

The multivariate test showed strong overall evidence of differences between intruder and control presentation ($p = 0.005$). This overall difference is the result of differences between control and extended conditions in Time to Reemerge and Time to Full Exposure.

There was evidence of period effects for Time to Full Exposure ($p = 0.0408$ in the second day, lizards re-emerged fully sooner, suggesting habituation). More importantly, for both Time

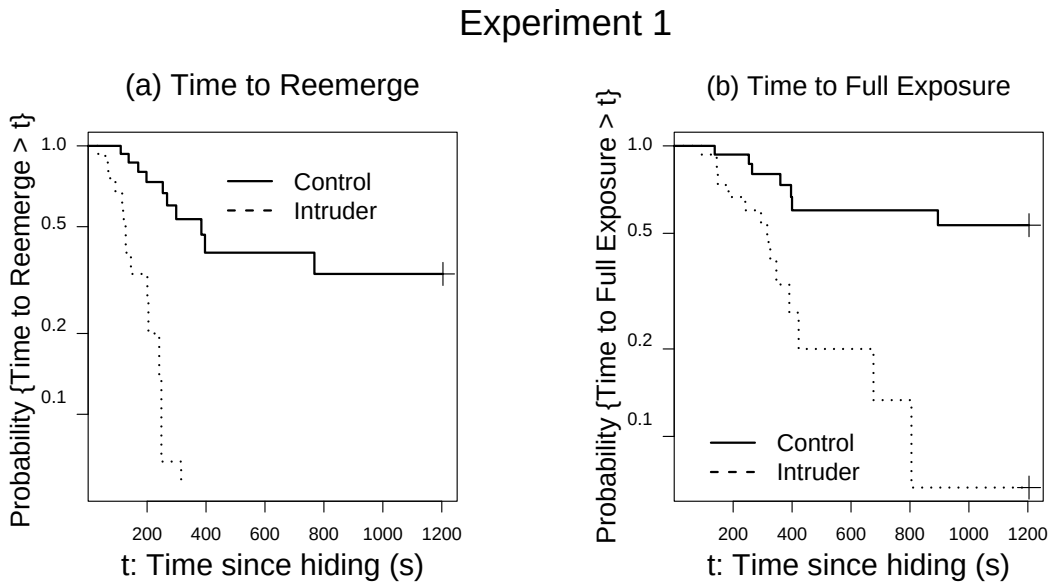

Figure 1.1: Experiment 1, (a) Time to Reemerge and (b) Time to Full Exposure. Survival curves (based on the Kaplan-Meier estimator of the survival function). The y-axis can be interpreted as (a) "Probability of not having re-emerged" and (b) "Probability of not having fully re-emerged." The cross denotes censoring. These figures do not take into account that measures for the same individual are potentially correlated and that there are two distinct sequences; they should not be used directly for hypothesis testing. P-values for treatment effects (analysis following Feingold & Gillespie, 1996) are 0.0025 and 0.0058, respectively.

to Reemerge and Time to Full Exposure, lizards re-emerged sooner if they had been in an aggressive encounter instead of given a control treatment (Fig. 1.1, p. 12; $p = 0.0025$ and 0.0058 for Time to Reemerge and Time to Full Exposure, respectively). Thus, the results for Time to Reemerge and Time to Full Exposure are consistent and in the direction predicted by the first hypothesis. Analyses using mixed-effects models yielded the same qualitative results. None of the analyses for any of the variables showed evidence of carry-over effects ($p > 0.4$).

There were no differences between control and extended treatment for (log of) Minimum Distance. For (square root of) Approach Distance I found a significant interaction between treatment and enclosure area ($F_{1,13} = 12.86$, $p = 0.0033$): Approach Distance increased with area in the control treatment, but not in the extended treatment (from a reparameterised model, regression coefficients for control and intruder are 1.03 and -0.385, respectively; s.e.= 0.414; $t_{18.6} = 2.48$ and -0.93, $p = 0.0227$ and 0.3654). There was weak evidence ($F_{1,12} = 4.51$, $p = 0.0552$) for a main effect of type of enclosure: approach distance was larger in enclosures with two females than in enclosures with one female and one small male (back-transformed least squares means are 7.4 and 4.11 m respectively). Although the speed of my approach did not differ between treatments (mean difference intruder-control (s.e. $= 0.018 \pm 0.021$ m/s, paired $t_{13} = -0.8675$, $p = 0.401$), I included my approach speed in the models for Approach Distance; neither the main effect nor its interaction with treatment were significant ($p > 0.3$).

1.4.2 Experiment 2: Differences between extended and immediate effects.

The multivariate test showed strong evidence of overall differences between extended and immediate effects ($p = 0.0130$). This overall difference was due to differences in Approach Distance and Minimum Distance.

Time to Reemerge and Time to Full Exposure did not differ between extended and immediate treatments. For (log of) Time to Full Exposure animals re-emerged sooner in later periods of testing: the final model included only a linear effect of period ($F_{1,33.2} = 12.41$, $p = 0.0013$;

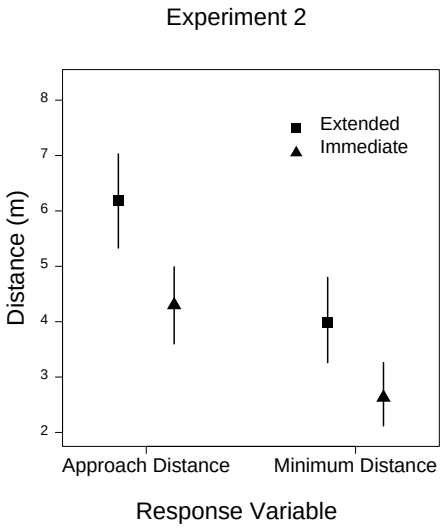

Figure 1.2: Experiment 2, Approach and Minimum Distance. Back-transformed adjusted means (one s.e. This figure should not be used for hypothesis testing. P-values for treatment effects (from mixed model) are 0.0240 and 0.0236 for Approach and Minimum distances, respectively.

regression coefficient ± s.e. = -0.254 (± 0.072), suggesting habituation. Analyses with Feingold & Gillespie's (1996) method also indicated no treatment effects.

Approach Distance and Minimum Distance differed between extended and immediate treatments. For (log of) Minimum Distance there were effects of both treatment and period; the final model included a quadratic term for Period ($F_{1,20.9} = 6.42$, $p = 0.0194$; coefficient for linear term = 0.401, coefficient for quadratic term = -0.123) and a term for treatment ($F_{1,4.81} = 10.68$, $p = 0.0236$). As period of testing progressed, Minimum Distance decreased, suggesting habituation; more importantly, Minimum Distance in the immediate treatment was shorter than in the extended treatment (Fig. 1.2, p. 14). For Approach Distance there was only an effect of treatment ($F_{1,30.2} = 5.65$, $p = 0.0240$). There was a 7% difference in my approach speed between treatments (mean speeds for extended and immediate were 0.442 m/s and 0.473 m/s, respectively; $F_{1,29.5} = 5.82$, $p = 0.0223$, from a mixed model using lizard as random effect). However, neither the interaction of approach speed with treatment, nor the main effect of approach speed had any significant effect on Approach Distance (interaction: $F_{1,17.6} = 1.04$, $p = 0.3216$; main

effect: $F_{1,8.42} = 0.7$, $p = 0.6143$). In summary, the results for both Minimum and Approach Distance are consistent and in the direction predicted by the second hypothesis: lizards allowed the potential predator to approach closer when they were engaged in an ongoing fight with a conspecific intruder (Fig. 1.2, p. 14).

A possible explanation of the differences in Approach and Minimum distances are dilution effects (see discussion). In Experiment 2 I also recorded whether the female was out of the refuge. If dilution effects are important, experimental lizards should show shorter Approach or Minimum distances when the female was out of the refuge. I compared the effect of a female out on Approach and Minimum distances for the extended treatment. I also reanalysed the final models for Approach Distance and Minimum Distance, allowing for the effect of female presence/absence to differ between treatments. In no case was the presence of the female significant (all $p > 0.15$).

No experiment compared immediate effects with a control. However, if we assume that the animals from Experiment 2 would have shown differences between extended and control in the same direction as animals from Experiment 1 did, we can summarise the results from both experiments together as shown in Fig. 1.3 (p. 16).

1.5 Discussion

Past aggressive interactions (Experiment 1) decreased the amount of time male *T. hispidus* spent hiding after a simulated predatory attack; when the predator attacked during an ongoing aggressive encounter (Experiment 2), lizards also allowed the predator to approach closer (Fig. 1.3, p. 16). These results show: a) the existence of extended effects of aggressive behaviour on antipredator behaviour; b) that extended effects differ from immediate ones. The results are consistent with the two economic (adaptive) hypotheses stated in the introduction: a) past presence of an intruder can indicate an increase in the probability of future intrusions,

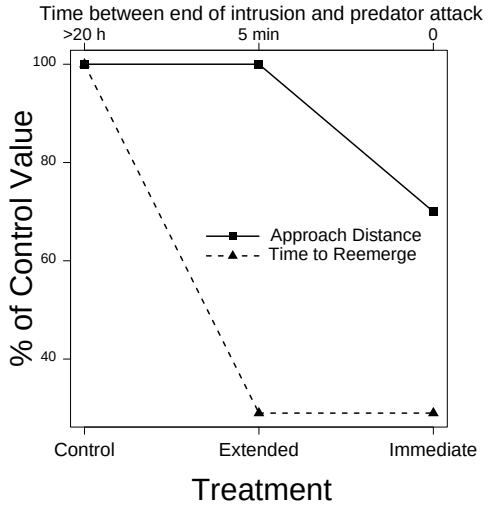

Figure 1.3: Summary of results from both experiments, based on Approach Distance and Time to Reemerge. I calculated "% of Control Value" as $\frac{100^{*}\text{ Adjusted mean for experimental condition}}{\text{Adjusted mean for control}}$.

and therefore if a predator attacks soon after an aggressive interaction is over, a territorial resident should modify its behaviour to decrease the chances of territorial intrusions at the cost of increased predation risks, and b) current presence of an intruder increases territorial costs of hiding with respect to past presence of an intruder, and thus territorial residents should show further increases in their exposure to predation when the predator approaches during an aggressive encounter.

Extended effects of aggression on antipredator behaviour (Experiment 1) have not been reported before, but the increase in predation exposure when the lizards were involved in a fight 5 min before the attack of the predator is consistent with economic models of antipredator behaviour (Ydenberg & Dill 1986; Clark 1994). The results indicate that extended effects affect mainly re-emergence time, not approach distances. A predatory attack is generally a fast event and the rate of increase of the ability to monitor the territory by delaying flight is probably small compared to the rate of increase of mortality risk. Thus, extended effects on approach distances are likely to be non-existent or difficult to detect when present. In contrast, changes

in re-emergence can result in increased ability to monitor the territory without large increases in mortality risk.

The immediate effects (Experiment 2) are consistent with those observed by Jakobsson et al. (1995) in both the cichlid, *Nannacara anomala*, and the warbler, *Phylloscopus trochilus*, where animals engaged in an aggressive interaction allow a predator to approach closer than animals exposed to a control stimulus (see also Brick, 1998). The data presented here also show that immediate effects resulted in a decrease in time to reemerge (with respect to a control). However, the immediate effects did not result in further decreases in times to reemerge compared to the extended effects, despite the potentially larger intrussion costs in the immediate condition (see Introduction).

In general we should expect different components of the antipredator behaviour to be differentially affected by aggressive interactions, as hiding quickly can have very different consequences in terms of mortality from predation and intruder detection than re-emerging late. These results emphasise the need of measuring the components of the antipredator strategy that best characterise the key behavioural decisions involved in predator avoidance (e.g., Lima & Dill 1990) and intruder detection.

The immediate effects on Approach and Minimum Distance (Experiment 2) could be explained by the non-adaptive "sensory limitation hypothesis:" an animal involved in a fight might be unable to detect a predator as fast as an animal that is not involved in a fight (e.g., Bernays & Wcislo 1994; Milinski 1984). Sensory limitation seems to be the mechanism invoked by Brick (1998) and by Jakobsson et al. (1995) to explain the decrease in approach distance during intraspecific fights in both warblers and cichlids. In its most extreme form, the sensory limitation hypothesis predicts that an animal will initiate escape as soon as the predator is detected. In contrast, the economic hypothesis emphasises the decision component (Ydenberg & Dill 1986): the decrease in approach distance in the immediate treatment would be the result of a change in the perceived cost of hiding and not of a decrease in the ability to detect predators. It is not

possible to differentiate between the two hypothesis with the approach distance data, as both make similar predictions regarding approach distance in the first approach of the predator. It is difficult to determine the exact moment when a predator is detected, but the two hypotheses could be differentiated by increasing the costs of hiding: the economic hypothesis would predict increased exposure to predation, whereas the sensory limitation hypothesis would predict no change in antipredator behaviour. Further work to elucidate whether the changes in approach distance in the immediate condition are due to sensory limitations, to an economic decision, or a combination of both, is warranted.

A third explanation for the reduction in approach distance in the immediate treatment are dilution effects: if the predator can only capture a single prey the chances that the resident is the victim decrease in the immediate treatment because there are two lizards in the area. The tests in Experiment 2 (presence vs. absence of female out of the refuge), although do not conclusively exclude dilution effects, suggest that the changes in approach and minimum distances in the immediate treatment were not solely a result of dilution effects.

In contrast, the differences in Time to Reemerge and Time to Full Exposure between the control and the extended conditions (Experiment 1) cannot be explained by the sensory limitation hypothesis or by dilution effects. Thus, the economic hypothesis provides the best explanation for the changes in time to reemerge.

Past aggressive interactions with intruders can affect the subsequent behaviour of a territorial holder. Great tits invest more time in territorial vigilance (at the cost of decreased foraging) after encountering intruders (Ydenberg & Krebs 1987; Kacelnick et al. 1981); in the lizard *Sceloporus jarrovi* the frequency of most displays' peaks shortly after an encounter (Moore 1987; also Thompson & Moore 1992 for *Urosaurus ornatus*); in several taxa, following a previous victory, there is an increase in the probability of winning subsequent encounters (Adamo & Hoy 1995; Chase et al. 1994). Functionally, these different phenomena can be a response by the territorial resident to a transient increase in the probability of re-intrusion by

the same intruder; and extended effects of aggression on antipredator behaviour are consistent with minimisation of the increased risk of territorial intrusion caused by a transient change in the probability of future intrusions. Thus, a similar functional explanation can underlie different behavioural phenomena where animals change their aggressive/antipredator behaviour as a response to local changes in their social environments (e.g., Oliveira et al., 1998).

Extended effects show a connection between antipredator and aggressive behaviour which should vary with the defensibility of resources, and that can influence the (co)evolution of these sets of traits, by increasing both predation related costs of territorial behaviour and territorial costs of hiding. The hypothesis underlying extended effects is testable, using both within- and among-species comparisons. Given that an economic reasoning is the basis of the extended effects, it will also be particularly important to understand the relative contributions of perceptual constraints, dilution effects, and increased hiding costs in the effects of an ongoing fight on approach distances, and, ultimately, measure the fitness consequences of different antipredator responses following an aggressive encounter.

1.6 Acknowledements

I thank C. Lázaro-Perea for initial discussion, subsequent frequent discussion and advice, encouragement, and help building, cleaning, and taking the enclosures down, as well as logistic and other types of support; Departamento de Fisiologia, Universidade Federal do Rio Grande do Norte, Brazil, and in particular M. F. Arruda for invaluable logistic support; Eflex-IBAMA of Nisia Floresta, specially J. Dantas, for facilities, tools, and permission to perform this work; E. Nascimento and several children of Porto and Nisia Floresta for assistance with lizard capture; C. A. Marler and C. T. Snowdon for advice; D. Abbott, A. Bakshy, J. K. Bester-Meredith, S. de la Torre, K. F. Klomberg, C. Lázaro-Perea, C. A. Marler, C. W. Meredith, E. V. Nordheim, C. T. Snowdon, and B. C. Trainor, and two anonymous reviewers for comments on the ms; E. V. Nordheim for statistical discussion; A. C. de Moura and S. Profirio for the first mealworms;

D. Bates, J. Pinheiro, and Z. Shkedy for answering questions about nlme. Manuscript preparation partially supported by a John and Virginia Emlen Graduate Student Fellowship, Dept. Zoology, University of Wisconsin-Madison.

1.7 References

Adamo, S. A. & Hoy, R. R. 1995. Agonistic behaviour in male and female field crickets, Grillus bimaculatus, and how behavioural context influences its expression. Animal Behaviour, 49, 1491-1501.

Arcese, P. 1987. Age, intrusion pressure and defence against floaters by territorial male song sparrows. Animal Behaviour, 35, 773-784.

Bernays, E. A. & Wcislo, W. T. 1994. Sensory capabilities, information processing, and resource specialization. Quarterly Review of Biology, 69, 187-204.

Brick, O. 1998. Fighting behaviour, vigilance and predation risk in the cichlid fish Nannacara anomala. Animal Behaviour, 56, 309-317.

Chase, I. D., Bartolomeo, C., & Dugatkin, L. A. 1994. Aggressive interactions and inter-contest interval: how long do winners keep winning? Animal Behaviour, 48, 393-400.

Clark, C. W. 1994. Antipredator behaviour and the asset-protection principle. Behavioural Ecology, 5, 159-170.

Diggle, P. J., Liang, K.-Y., & Zeger, S. L. 1994. Analysis of Longitudinal Data. Oxford: Oxford University Press.

Edgington, E. S. 1995. Randomization Tests, 3rd ed. New York: Marcel Dekker.

Feingold, M. & Gillespie, B. W. 1996. Cross-Over Trials With Censored Data. Statistics in Medicine, 15, 953-967.

Jakobsson, S., Brick, O., & Kullberg, C. 1995. Escalated fighting behaviour incurs increased predation risk. Animal Behaviour, 49, 235-239.

Johnson, W. D. & Mercante, D. E. 1996. Analysing multivariate data in crossover designs using permutation tests. Journal of Biopharmaceutical Statistics, 6, 327-42.

Jones, B. & Kenward, M. G. 1989. Design and Analysis of Cross-Over Trials. New York: Chapman & Hall.

Kacelnick, A., Houston, A. I., & Krebs, J. R. 1981. Optimal foraging and territorial defence in the great tit (Parus major). Behav. Ecol. Sociobiol., 8, 35-40.

Lawless, J. F. 1982. Statistical models and methods for lifetime data. New York: John Wiley & Sons.

Lima, S. L. 1998. Stress and decision making under the risk of predation: recent developments from behavioural, reproductive, and ecological perspectives. In: Advances in the study of behaviour, v. 27. (Ed. by P. J. B. Slater, J. S. Rosenblat, C. T. Snowdon, & M. Milinski), pp. 215-290. San Diego: Academic Press

Lima, S. L. & Dill, L. 1990. Behavioural decisions made under the risk of predation: a review and prospectus. Candian Journal of Zoology, 68, 619-640.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. 1996. SAS System for Mixed Models. Cary, NC: SAS Institute.

Milinski, M. 1984. A predator's cost of overcoming the confusion-effect of swarming prey. Animal Behaviour, 32, 1157-1162.

Moore, M. C. 1987. Circulating steroid hormones during rapid aggressive responses of territorial male mountain spiny lizards, Sceloporus jarrovi. Hormones and Behaviour, 21, 511-521.

Oliveria, R. F., McGregor, P. K., & Latruffe, C. 1998. Know thine enemy: fighting fish gather information from observing conspecific interactions. Proc. R. Soc. Lond. B., 265, 1045-1049.

Pinheiro, J. C & Bates, D. M. 2000. Mixed-Effects Models in S. Springer-Verlag, in press.

Rodrigues, M. T. 1987. Sistemática, ecologia e zoogeografia dos Tropidurus do grupo Torquatus ao sul do rio Amazonas (Sauria, Iguanidae). Arquivos De Zoologia, S. Paulo, 31, 105-230.

Stamps, J. A. & Krishnan, V. V. 1995. Territory acquisition in lizards: III. Competing for space. Animal Behaviour, 49, 679-693.

Stamps. J. A. & Krishnan, V. V. 1998. Territory acquisition in lizards. IV. Obtaining high status and exclusive home ranges. Animal Behaviour, 55, 461-472.

Statistical Sciences. 1995. S-Plus guide to statistical and mathematical analysis, version 3.3. StatSci, Seattle.

Stutchbury, B. J. 1991. Floater behaviour and territory acquisition in male purple martins. Animal Behaviour, 42, 435-443.

Thompson, C. W. & Moore, M. C. 1992. Behavioural and hormonal correlates of alternative reproductive strategies in a polyginous lizard: tests of the relative plasticity and challenge hypotheses. Hormones and Behav., 26, 565-585

Vitt, L. J. 1995. The ecology of tropical lizards in the caatinga of northeast Brazil. Occasional Papers of the Oklahoma Museum of Natural History, 1, 1-29.

Yasukawa, K. 1979. Territory establishment in red-winged blackbirds: importance of aggressive behaviour and experience. Condor, 81, 258-264.

Ydenberg, R. C. & Dill, L. M. 1986. The economics of fleeing from predators. In: Advances in the study of behaviour, v. 16. (Ed. by J. S. Rosenblatt, C. Beer, M.-C. Busnel, & P. J. B. Slater), pp. 229-249. New York: Academic Press

Ydenberg, R. C. & Krebs, J. R. 1987. The tradeoff between territorial defence and foraging in the great tit (Parus major). American Zoologist, 27, 337-346.

Chapter 2

Antipredator behavior changes following an aggressive encounter: effects of testosterone manipulations

2.1 Abstract

Changes in antipredator behavior following a conspecific territorial invasion suggest that males of the lizard *Tropidurus hispidus* incur greater predation risks to minimize potential costs from future conspecific intrusions. Elevated testosterone levels in several lizard species result in males that increase their allocation to territorial defense at the expense of other costs. Consequently, we expected that elevated testosterone would: (1) increase exposure to predation; (2) produce a disproportionate increase in exposure to predation following a past aggressive interaction. We manipulated testosterone levels of male *T. hispidus* using subcutaneous testosterone implants. Our results provide strong evidence that past aggressive interactions result in increased exposure to predation and that the type of first encounter (aggressive interaction with a conspecific vs. control presentation) had long-lasting effects on antipredator behavior. We found no evidence of an association between aggressive and antipredator behaviors, or of differences in aggressive behavior related to testosterone treatment (but we found indication of a decrease in aggression with increasing corticosterone plasma levels). There was evidence of changes in antipredator behavior associated to hormonal treatments, but in a different direction from the

one hypothesized. We discuss these results in the context of absence of changes in aggressive behavior related to testosterone, possibly related to housing conditions, and suggest that future studies might benefit from focusing on the role of corticosterone.

2.2 Introduction

Recent empirical work (Díaz-Uriarte, 1999) has shown that past aggressive interactions can affect antipredator behavior: males of the lizard *Tropidurus hispidus* reemerge sooner from a refuge after hiding from a predator if the predator attacks 5 min after the resident male has chased away a conspecific intruder male. Those results are consistent with the economic hypothesis that male *T. hispidus* incur greater predation risk to minimize the potential cost of future territorial intrusions. Functionally, this hypothesis rests on two assumptions: first, that intruders are persistent, so that a past aggressive interaction can inform a territorial resident that subsequent intrusions are likely; second, that successful defense of a territory is relevant for reproductive success. The territorial costs of hiding increase if intruders are persistent because the same amount of time in hiding can result in a much larger decrease in reproductive success if an intruder is likely to return.

The proximate mechanisms underlying the change in antipredator behavior following an aggressive encounter are unknown. There is evidence from sheep (Vandenheede & Bouissou, 1996; Bouissou & Vandenheede, 1996) that increased testosterone results in decreased fearfulness (testosterone can also decrease nest defense against predators by decreasing the likelihood of males with high testosterone being present at a nest —Cawthorn et al., 1998—, but this is not contradictory with the previous evidence). Other hormones such as thyroid hormone and growth hormone (e.g., Abrahams & Pratt, 2000; Abrahams & Sutterlin, 1999; Johnsson et al., 1999, 1996) have been shown to affect antipredator responses by modifying the antipredator-foraging trade-off. To our knowledge, there is no previous work on the effects of testosterone on antipredator behavior operating by modifying, physiologically, the antipredator-territorial

defense trade-off. However, it has been well documented that androgen hormones are involved in territorial defense in lizards (e.g., *Sceloporus jarrovi*: Moore, 1988; Moore & Marler, 1987; Marler & Moore, 1989, 1991; *Anolis sagrei*: Tokarz, 1987, 1995; *Uta stansburiana*: DeNardo & Sinervo, 1994; *Psammodromus algirus*: Salvador et al., 1997). Therefore, the evidence from temperate-zone lizards indicates that increased testosterone levels result in males that increase their allocation to territorial defense at the expense of other costs (such as survivorship or foraging).

If testosterone does increase the allocation to territorial defense, we would anticipate that increases in testosterone will modify the trade-off between antipredator behavior and territorial defense so that, when faced with a predator, animals with increased testosterone levels will incur greater predation risks to minimize the risk of territorial loses. This effect should manifest itself as a decrease in the distance between predator and prey when the prey initiates escape (approach distance) and/or a decrease in the time to reemerge from a refuge following a predatory attack. In addition, the effects of testosterone on antipredator behavior could be enhanced if the predator attacks shortly after a territorial resident has evicted a conspecific intruder because the territorial costs of hiding can be particularly high. Thus, elevated testosterone levels are predicted to cause a disproportionate decrease in approach distance and/or time to reemerge when the predator attacks shortly after a conspecific intrusion. In other words, we should expect an interaction between testosterone level and the effects of a past territorial intrusion on antipredator behavior. These hypotheses can be examined using hormonal manipulations.

Manipulation of hormone levels can also help investigate whether aggression and antipredator behavior are physiologically linked, and can increase the variation in aggressive and antipredator responses thus making covariation patterns among these two sets of traits more detectable (e.g., see Sinervo & Basolo, 1996 for discussion of phenotypic manipulations). Genetic correlations among different functional categories of behavior could have dramatic effects on behavioral evolution, because of correlated responses to selection (Stamps, 1991). The si-

multaneous collection of data on aggression and antipredator behavior allow one to examine if, at least phenotypically, these two sets of behaviors are correlated. Despite the value of this approach, studies that focus on the correlation of functional categories with major fitness effects are still rare (Sih, 1992; Stamps, 1991). Nevertheless, a phenotypic correlation between aggressive and antipredator behavior has been found in a few cases (spiders: Reichert & Hedrick, 1993; see also Huntingford, 1976; Tulley & Huntingford, 1988), and it has been suggested that hormones could be the link between these two functional categories (Reichert & Hedrick, 1993; Stamps, 1991). But this hypothesis has not been tested.

In this paper we examine the effects of testosterone manipulations on antipredator behavior and on the changes in antipredator behavior following a conspecific intrusion, in males of the lizard *Tropidurus hispidus*, for which there is evidence that past aggressive interactions result in changes in antipredator behavior (Díaz-Uriarte, 1999). We also present data on testosterone plasma levels and the effects of testosterone manipulations on the aggressive behavior of a tropical species of lizard; most of the evidence for the effects of testosterone on aggressive behavior of lizards comes from temperate-zone species (see references above). In contrast, both male and female *Tropidurus hispidus* are aggressive and territorial (pers. obs.), and are capable of reproducing throughout the year (pers. obs.; also VanSluys, 1993).

2.3 Methods

2.3.1 Animals and study site

Experiments were conducted between 26 July 1997 and 2 January 1998 at the Nisia Floresta Forest Experimental Station, EFLEX-IBAMA, (6° 5' S, 35° 12' W), located 45 km from Natal (Northeastern Brazil). We used adult males of the lizard *Tropidurus hispidus* (snout-vent length [SVL] 70-130 mm), a widespread, diurnal, sit-and-wait iguanine lizard in South America (Rodrigues 1987; Vitt 1995). Experimental subjects were adult males (SVL $\geq$ 100 mm) captured

in villages close to the station that had not been used in other experiments. Intruders (adult males SVL $>$ 90 mm) were used a maximum of three times and were never injured by the experimental procedure. The same experimental animal was not exposed to the same intruder more than once. All animals were released in the area of capture at the end of testing.

2.3.2 Enclosures and animal husbandry

Details on enclosures and animal husbandry are described in Díaz-Uriarte (1999). Briefly, we used enclosures to minimize variation in behavior. Enclosures were located in open patches in plantation areas and measured 2 by 2 m. Enclosures were 1 m high, constructed from transparent plastic attached to a wood frame. Each enclosure contained two refuges made with bricks and roof tiles that offered protection and were readily used by the lizards as hiding places. Enclosures were partially covered from above to provide shade during the central hours of the day, and also included one adult female. In four cases, females disappeared before the end of the testing period, probably from predation. All females were randomly assigned to enclosures/males, except that females SVL had to be at least 5 mm less than that of the males (in the field, males were associated with females smaller than themselves; pers. obs.).

We placed a blind 7.5 m away from the enclosure. Using suspended fishing lines, we could move an intruder from behind the blind to inside the enclosure and retrieve it at the end of the trial, without ever leaving the blind. Enclosures were more than 15 m apart with dense and tall intervening vegetation ensuring no visual contact between them, and were placed in areas where, during a period of ten months, we only observed four free-ranging adult *T. hispidus* (one male, three females). Thus, interactions with naturally-occurring conspecifics should have been extremely rare.

Every two to three days lizards were fed a diet of crickets, meal worms, fly maggots, roaches and beetles (dusted with a multi-vitamin preparation —Reptivite— once a week), and a mixture of egg, powdered milk, and fruit, and were provided with water in several water containers.

Between trials, we thoroughly cleaned all bricks and tiles and either removed the upper 3-5 cm of soil, or added 3-5 cm of soil, to minimize the persistence of possible chemical marks or pathogens from previous residents. Animals in the enclosures displayed normal antipredator, aggressive, and mating behavior (see Díaz-Uriarte, 1999). While in the enclosures, lizards were rarely approached by humans.

2.3.3 Experimental design, antipredator tests, aggressive behavior

This study involved two experimental factors: an among-individual treatment (hormonal treatment) and a within-individual treatment (territorial intrusion). The hormonal treatment had three levels: empty implant (control), single testosterone implant, double testosterone implant. At the time this experiment was carried out, no information was available on the natural range of variation of testosterone levels in this species and therefore we used three different levels for the testosterone manipulation. In the empty group, animals were given two empty implants; in the single implant, animals were given one empty and one testosterone-filled implant, and in the double group, animals were given two testosterone-filled implants. Hormone implants were of silastic tubing (5 mm packed length; ID 1.47 mm, OD 1.96 mm) and were placed subcutaneously, one in each side of the body, after making a small incision. Implants had been left in saline solution for 24 h before implantation. Before surgery, animals were immobilized with cold and given lidocaine (0.02 ml, 0.2% solution) in the place of the incision. By the time the lizards were released back in the enclosures they were fully active. The implants used in this study were of the same size as those used for male *Sceloporus jarrovi* (Marler & Moore, 1988), which weigh approximately half as much as *T. hispidus* adult males.

The territorial intrusion treatment (the within-individual treatment) had two levels: intruder and control. In both cases, we presented the male lizard with a stimulus (intruder or control) and, five minutes later, measured its antipredator behavior by simulating a predatory attack. A test (stimulus presentation + antipredator test) lasted approximately 40 min per

animal. In the intruder condition, we introduced an intruder adult male and left it inside the enclosure for a maximum of 15 min. Once the resident attacked, the intruder was left inside for 3 min or six attacks, whichever came first (this is slightly different from Díaz-Uriarte, 1999). If the resident had not directed at least three attacks during the 3 min, the intruder was left inside the enclosure until that criterion was met. After the trial was over, we retrieved the intruder to the blind, and then waited another 2 min before carrying out the antipredator test. In the control condition, we introduced a wood stick (of approximately the same size and color as an adult male) in place of an intruder, and left it inside the enclosure for 4 min 10 sec (the median latency to attack from data in Díaz-Uriarte, 1999).

The variables used to characterize antipredator behavior (Table 2.1, p. 30) reflect the two key behavioral decisions of a *T. hispidus* faced with an attacking predator: when to initiate escape from the predator and, after hiding, when to reemerge from the refuge. We used a human as a simulated predator. To run the antipredator test, one of us positioned himself 13 m away from the enclosure (4.5 m behind the blind) and approached the lizard directly at a moderate speed (mean = 0.22 m/s, s.d. = 0.036 m/s). Whenever the lizard moved, the experimenter stopped for 15 sec and recorded his position and the lizard's position, and then approached again. The approach-and-stop continued until the lizard hid, and then the experimenter moved to a spot 4.5 m from the enclosure, and remained motionless for 20 min. The experimenter recorded all his own movements and the lizard's behavior using an HP-48GX calculator for continuous event recording. All tests and observations were conducted by the same person (R. D.-U.) when lizards were active and air temperature (shaded bulb at 1.5 m) was higher than 26 °C. The experiment was blind with respect to hormone treatment: when antipredator tests were conducted, the experimenter was unaware of the hormone treatment group of the lizards. Aggressive behavior was characterized using the four variables shown in Table 2.2 (p. 30), measured during the presentation of the intruder.

The territorial intrusion treatment was applied according to a typical cross-over trial (e.g.,

Table 2.1: Response variables used to measure antipredator behavior.

Variable	Description
Approach distance	Distance between observer and the lizard when the lizard first initiated flight.
Minimum distance	Minimum distance between the observer and the lizard before it initiated flight; the same as approach distance if lizards run directly to hiding or hide within 15 sec of their first flight.
Time to reemerge	Time since the lizard hid until it reemerged (i.e., until at least all the head was visible out of the refuge).
Time to full exposure	Time since the lizard hid until it was fully exposed (all the lateral surface of the body —not including the tail— was visible out of the refuge.

Table 2.2: Response variables used to measure aggressive behavior.

Variable	Description
Latency to first attack	Time between when the intruder is introduced in the enclosure and the resident directs its first attack (rapid movement towards the intruder) or bite.
Interval between first and third attacks	Time between when the resident directs its first and the third attack or bite.
Displays before attack	Number of displays (head bobs, push-ups) by the resident since the intruder is introduced in the enclosure until the resident directs its first attack.
Displays after attack	Number of displays (head bobs, push-ups) by the resident between the time the intruder is returned behind the blind and the antipredator test is started. This is a 2 min period. The time it takes to return an intruder behind the blind is approximately 2 min.

Jones & Kenward, 1989; Díaz-Uriarte, 2000 a & b); we used the two sequences CIIC and ICCI (i.e., animals in sequence CIIC were first given the control treatment, the following day the intruder treatment, the third day the intruder and the fourth day the control treatment). Therefore, the experimental unit is different for the within- and the among-individual treatments. The effects of a territorial intrusion are estimated using within-animal comparisons, whereas the effects of the hormonal manipulation are estimated using among-animal comparisons.

In addition to the two experimental treatments, two potential sources of variation are enclosures and batches. The same six enclosures were used throughout the experiment. We conducted trials in batches, with each batch containing six males (one per enclosure). We used batches as a form of blocking because: 1) we had no information about possible variation in testosterone levels throughout the year, and this experiment was run over a six month period; 2) there was temporal variation in the type of food available; using batches we could ensure that, within a batch, all animals were provided with the same type and quantity of food, and at similar days/hours. In each batch, two males were assigned to each of the three levels of the hormone treatment. Within each of the hormonal treatments, one male was randomly assigned to one of the sequences and the other male to the other sequence. Assignment of animals to the hormonal treatment was by restricted randomization (certain assignments were not allowed). Animals were ranked by mass and randomized among hormonal treatments. Non-allowed combinations were those where the same hormonal treatment would have been assigned to either the two largest or the two smallest animals. Thus, out of a total of 90 possible assignments, 30 were not allowed. This was done to ensure adequate interspersion (Hurlbert, 1984) with respect to size to eliminate the possibility of confounding hormonal manipulations with variations in size.

The experiment was designed so that, at completion, each enclosure would have been used twice with each hormonal treatment (and once with each combination of hormonal treatment by sequence). This results in a layout resembling a Latin square: in each batch, the three hormonal

Table 2.3: Experimental design: enclosure, batch, hormonal treatment, and aggression-treatment sequence. D: double testosterone implant. S: single testosterone implant; E: two empty implants. Bold: sequence CIIC for the aggression treatment. Empty cells are missing data. In two cases (one S-male and one E-male), the animals could not be tested because of extreme shyness and lack of habituation. In three cases (two S-males and one E-male) males died during the study. In two cases (one S-male and one E-male) males disappeared, probably because they were eaten by an opossum (*Didelphis albiventris*). One animal (S-male) could not be tested because, during the three week period, there was continuous human activity around the enclosure. Note that the pattern of missing data cannot be related to the hormonal manipulation.

		Enclosure					
		1	2	3	4	5	6
B	1	**D**	**S**		S	**E**	D
a	2	**S**	E		**E**	D	**D**
t	3	E	S	**E**	D	**D**	**S**
c	4	S		D	**D**		E
h	5	**E**	D	**D**		E	S
	6	S	**E**				S

treatments are replicated twice, and for each enclosure over the whole experiment, the three hormonal treatments are replicated twice. We randomly chose the square used. This scheme was maintained for the first five batches. At the end of the fifth batch, however, we had been able to get behavioral measurements from ten doubly-implanted males, seven single-implanted males, and eight empty-implanted males. Keeping the same design for the sixth batch could have resulted in even further unbalance, and thus for the sixth batch we assigned three of the enclosures to single-implanted males and three to empty-implanted males, randomly. The actual design used is shown in Table 2.3 (p. 32).

In each batch the protocol was as follows. A male and a female were introduced in each enclosure on day one. They were fed and allowed to habituate for two or three days. On day four or five or early on day six, we took males out of the enclosures and surgically implanted them with testosterone-filled or empty silastic implants (see below). Males were returned to their enclosures within three hours. Thus, by day six all animals had been given hormone implants. For another 11 to 14 days (most studies with lizards that involve hormone implants

leave implants in place between one and three weeks before behavioral testing —e.g., Marler & Moore, 1989, 1991; DeNardo & Sinervo, 1994) animals were fed regularly and habituated to the intruder-delivery system (see Díaz-Uriarte, 1999 for details of habituation). During the next four days we measured antipredator behavior, as specified by the sequences of within-individual treatments. Most batches were completed by day 22. During the first hours of activity on day 23 (i.e., one day after the last test was completed), we entered the enclosures and obtained a blood sample from the males. All blood samples were obtained within 4 min of entering the enclosure, and the samples for all males in a batch were obtained within one hour. Usually a new batch of lizards was introduced in the enclosures on day 23 or 24. We removed the testosterone implants from males, and males and females were released in the areas where they had been captured. Animals were marked by toe-cliping; this allowed to individually identify each animal and prevented using the same animals more than once in the experiment. All animals within a batch were subject to the main manipulations (introducing them in enclosures, baseline tests, surgery) at the same time, but there were minor variations from batch to batch (because of weather). Throughout the study period the hours of testing changed to accommodate shifts in activity periods, and as summer progressed we also increased the shaded area within the enclosures.

Blood samples for hormone assays were collected from the post-orbital sinus using heparinized tubes. Blood was centrifuged, and plasma extracted and frozen at -10 °C. In addition to the experimental animals, during the months of January, March, and December we collected blood samples from another 33 adult males from several nearby areas (see Fig. 2.1, p. 39). To determine plasma levels of testosterone and corticosterone, radioimmunoasay was performed as described in Moore (1986) and Foufopoulos et al. (2000), following ether extraction of plasma and chromatographic separation of the steroid hormones from each other and from interfering lipids on a diatomaceous earth : propanediol : ethylene glycol microcolumns. Each sample was assayed in duplicate. Intra-assay coefficients of variation were 1.3% for testosterone and 1.8% for corticosterone. Inter-assay coefficients of variation were 4.6% for testosterone and 11.4% for

corticosterone. Testosterone and corticosterone plasma levels for all the experimental animals were determined in a single assay.

2.3.4 Statistical analyses

Effects of hormonal manipulations on hormone levels (testosterone and corticosterone) and aggressive behavior (Table 2.2, p. 30) were initially examined using linear mixed-effects models, with

$$y_{ijkl} = \mu + \eta_i + \xi_j + (\eta\xi)_{ij} + c_k + b_l + e_{ijkl}, \tag{2.1}$$

where y is the response, in the fixed effects part μ is the intercept, η is the hormone treatment, ξ is the effect of sequence, and the term in parentheses is the interaction hormone*sequence. In the random effects part, c, b, and e are the random effects of enclosure, batch, and individual, respectively; all the random effect terms are assumed normal and independent of each other. For hormone levels, only one measure per individual was available. For aggressive behavior, two observations were available; however, as the objective was to relate aggressive behavior to hormone treatment, before the analyses we obtained the mean of the two responses (or the mean of a suitable function of the responses, such as the log) for each individual. When testing for effects of batch and/or enclosure on testosterone and corticosterone plasma levels, however, the p-values were obtained from ANOVA models with batch and enclosure as fixed effects, since likelihood ratio tests of the hypothesis that a variance component is zero can be overly conservative (Pinheiro & Bates, 2000; Verbeke & Molenberghs, 1997).

We examined effects of hormonal manipulations on aggressive behavior with multivariate analysis of variance (MANOVA; Krzanowski, 1990; Morrison, 1990), with hormone treatment group as the explanatory variable and the four aggressive behavior variables as responses. We also examined the effects of plasma levels of testosterone and corticosterone and their possible interaction on aggressive behavior using multivariate regression (the extension of MANOVA

for continuous explanatory variables). Because of the exploratory nature of this part of the study, and to prevent for decreases in power related to violations of assumptions of MANOVA (equality of covariance matrices across groups) we also examined test-wise p-values of each of the responses variables. To provide protection against inflated Type I error rates, we adjusted for for multiple tests using Holm's sequentially rejective procedure (see Rice, 1989; Wright, 1992), with an family-wise error rate of 15 % (see Chandler, 1995).

Approach distance and minimum distance (Table 2.1, p. 30) were analyzed with linear mixed-effects models. We used the parameterizations in Jones & Kenward (1989), adding several covariates and random effects (see also Díaz-Uriarte, 2000 b); the full model examined was

$$y_{ijklmn} = \mu + \eta_i + \xi_j + (\eta\xi)_{ij} + c_k + b_l + s_{kl} + \pi_m + \tau_{n[j,m]} + \lambda_{n[j,m-1]} + (\eta\pi)_{im} + (\eta\tau)_{in} + (\eta\lambda)_{in} + e_{ijklmn}, \quad (2.2)$$

where everything is as in expression (2.1), with the addition of: π (period effect), τ (territorial intrusion effect), λ (first-order carry-over effect), and s (random effect of subject –lizard); terms in parentheses denote interactions. As is common in cross-over designs, we assumed only first order carry-over effects and no interactions of carry-over by treatment (i.e., carry-over of treatment A on treatment B is the same as carry-over of treatment A on treatment A). When analyzing approach distance, we also included a main effect for approach speed during the simulated predatory attack, as well as the interactions of approach speed with hormone treatment, sequence, and territorial intrusion. To examine the effects of plasma levels of testosterone and corticosterone, we used a model similar to (2.2), but we fitted simultaneously log testosterone and corticosterone plasma levels instead of hormone treatment. To account for possible non-linear effects of plasma levels of testosterone and corticosterone, we fitted models with quadratic terms and used added-variable plots (e.g., Hocking, 1996).

To fit the mixed-effects models, we proceeded as explained in Pinheiro & Bates (2000), Diggle

et al. (1994), and Littell et al. (1996). Briefly, we started with the full model, examining the fit of different covariance structures (compound symmetric, autoregressive, heteroscedastic) for the appropriate random effects; we used residual plots to asses the adequacy of the model, the need for transformations of the response, and possible influential points. After selecting a covariance structure, fixed effects terms were dropped sequentially from the model until all remaining terms had $p < 0.05$. If period (as categorical variable) was left in the model, we attempted to simplify this model by fitting linear and quadratic terms of period as a continuous variable. In addition, if the final model did not include some of the variables of primary importance (hormone treatment, territorial intrusion, their interactions, and the interaction of hormone treatment with sequence) we reexamined if they needed to be included in the final model.

The variables time to full exposure and time to reemerge had 46 and 11 out of 110 observations (about 42% and 10%) right-censored (i.e., in 46 trials lizards had not fully reemerged and in 11 trials lizards had not reemerged at the end of the 20 min observation period), and therefore require the use of survival analysis. We used Cox's proportional hazards model (e.g., Klein & Moeschbereger, 1997), with a full model analogous to the one in (2.2). Briefly, with this model the response is the hazard ratio, which can be thought of as the instantaneous probability of reemergence— given no reemergence until that moment; this hazard ratio is modeled as the product of a baseline hazard ratio*exponential of the sum of the covariate effects. To account for repeated measures within individuals, we used gamma frailty models (Klein & Moeschberger, 1997; Therneau & Grambsch, 2000; a frailty is equivalent to a random effect); these models generally yielded the same results as the marginal Cox model for multivariate survival data of Lee, Wei and collaborators (Lee et al., 1992; Lin, 1994; Wei et al., 1989). However, statistical tools for the inclusion of more than one frailty term are still not well developed; thus, for the survival analyses we regarded batch and enclosure as fixed-effects. We examined residuals for model adequacy and influential points (Klein & Moeschberger, 1997; Collet, 1994); model fitting proceeded analogous to approach and minimum distance (except testing was based on likelihood-ratio tests for frailty models).

In terms of the hypothesis discussed in the introduction, if testosterone modifies the effects of past aggressive interactions on antipredator behavior, we should observe a significant interaction term between hormone treatment and territorial intrusion treatment ($\eta\tau$). In the absence of this interaction, an overall change in antipredator behavior related to hormonal manipulations will be manifested as a significant main effect of hormone treatment (η).

We examined possible phenotypic correlations among antipredator and aggressive responses with canonical correlation analysis (e.g., Krzanowski, 1990; Morrison, 1990) using the four aggressive responses and two of the antipredator responses. Briefly, canonical correlation analysis attempts to find the largest possible correlation(s) between a linear combination of the first set of variables and a linear combination of the second set of variables; these linear combinations are the canonical variates, and the correlations among them are the canonical correlations. In our particular case, there are two canonical correlations (where the second canonical correlation is the largest possible, constrained by the new canonical variates being uncorrelated with the first ones). Thus, canonical correlation is somewhat similar to multiple regression, except both the "response" and the "predictors" are multivariate, and we make no distinction between response and predictor variables. We averaged, for each individual, the value of each aggressive response (or a suitable function of it, such as log) over the two aggressive encounters. For the antipredator responses, however, we only used the first trial where the animal had been subjected to a control (i.e., not a territorial intrusion); using all four trials for antipredator response could have confounded variation in antipredator behavior with variation in antipredator behavior following a territorial intrusion. Similarly, because of sequence and period effects in antipredator responses (see Results), the use of both control trials could have increased the variability of the responses, and it is not clear how to adjust for sequence effects in the presence of sequence*hormone interactions (see Results). For all observations from the first control trial Approach and Minimum distance had identical values, and thus only one of them was used in this analyses. Time to Full Exposure was not included as 40% of the observations were censored (see Results). We examined the hypothesis of no association between aggressive and

antipredator behavior by testing that the canonical correlations are zero with a likelihood ratio test, as explained in Krzanowski (1990, p. 447 and ff.).

Linear mixed-effects models were fitted using the R library nlme (Pinheiro & Bates, 2000) and SAS's PROC MIXED (Littell et al. 1996). Survival models were fitted with the survival 5 library (originally by T. Therneau, ported to R by T. Lumley) for R. Canonical correlations were performed with R (library mva). All p-values are two-sided.

2.4 Results

2.4.1 Effects of hormonal manipulations on hormone plasma levels and aggressive behavior

Figure 2.1 (p. 39) shows the plasma testosterone and corticosterone levels for the three treatment groups and a set of 33 wild adult males. We used (natural) log transformed data; analyses with data in the original scale showed apparent outliers and very highly influential points, as well as asymmetric normal probability plots; moreover, a log transformation helped stabilize the variance and might be a natural transformation for a measure of concentration (where, in the original scale, the variance can increase with the mean). For log testosterone, there was no evidence of either batch —i.e., seasonal changes— or enclosure effects ($F_{5,15} = 1.44$, $p = 0.2671$ and $F_{5,15} = 0.40$, $p = 0.8414$, respectively), but strong evidence ($F_{2,25} = 8.58$, $p = 0.0015$) of hormone treatment; these conclusions do not change if we exclude the individual from the single implanted group with lowest testosterone level (this individual had a studentized residual of -3.26, which is significant at the 0.05 level after bonferroni correction). For log corticosterone plasma levels there was no evidence of batch, enclosure, or hormone treatment effects ($F_{5,15} = 0.64$, $p = 0.6716$, $F_{5,15} = 2.00$, $p = 0.1362$, and $F_{2,25} = 1.33$, $p = 0.2833$, respectively). However, there was strong evidence of a decrease in the variance of log corticosterone plasma levels with hormone treatment ($\chi^2_2 = 10.63$, $p = 0.0049$ from a likelihood ratio test between

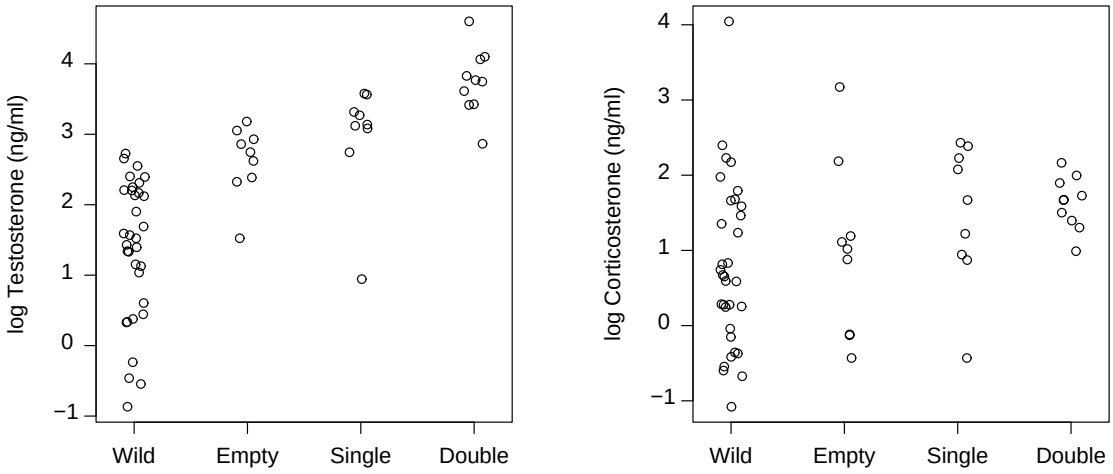

Figure 2.1: Plasma testosterone and corticosterone levels of the experimental animals and a set of 33 wild caught males.

heteroscedastic and homoscedastic models). There was no evidence of an interaction between sequence and hormone treatment or of a main effect of sequence in either corticosterone or testosterone plasma levels (interaction: $F_{2,22} = 0.41$ and 0.80, $p = 0.6692$ and 0.4642 for corticosterone and testosterone, respectively; main effect of sequence: $F_{1,22} = 0.23$ and 0.37, $p = 0.6363$ and 0.5490). There was no evidence of a difference in log corticosterone plasma levels between wild and empty implanted animals ($t_{15.03} = 0.59$, $p = 0.5648$ from a Welch two-sample t-test), but there was strong evidence of higher testosterone plasma levels in empty implanted than wild animals ($t_{17.2} = 3.23$, $p = 0.0049$), in spite of the large overlap in values. These results do not change if we only use wild animals with SVL > 100 mm ($t_{27.11} = 4.59$, $p < 0.0001$). There was no correlation between log plasma levels of testosterone and corticosterone ($\rho = 0.03$, $28\,d.f.$, $p = 0.4435$). Whether animals had been involved in an aggressive interaction the day before or two days before did not affect plasma levels of log testosterone ($F_{1,23} = 0.88$, $p = 0.3584$) or log corticosterone ($F_{1,26} = 0.84$, $p = 0.3666$).

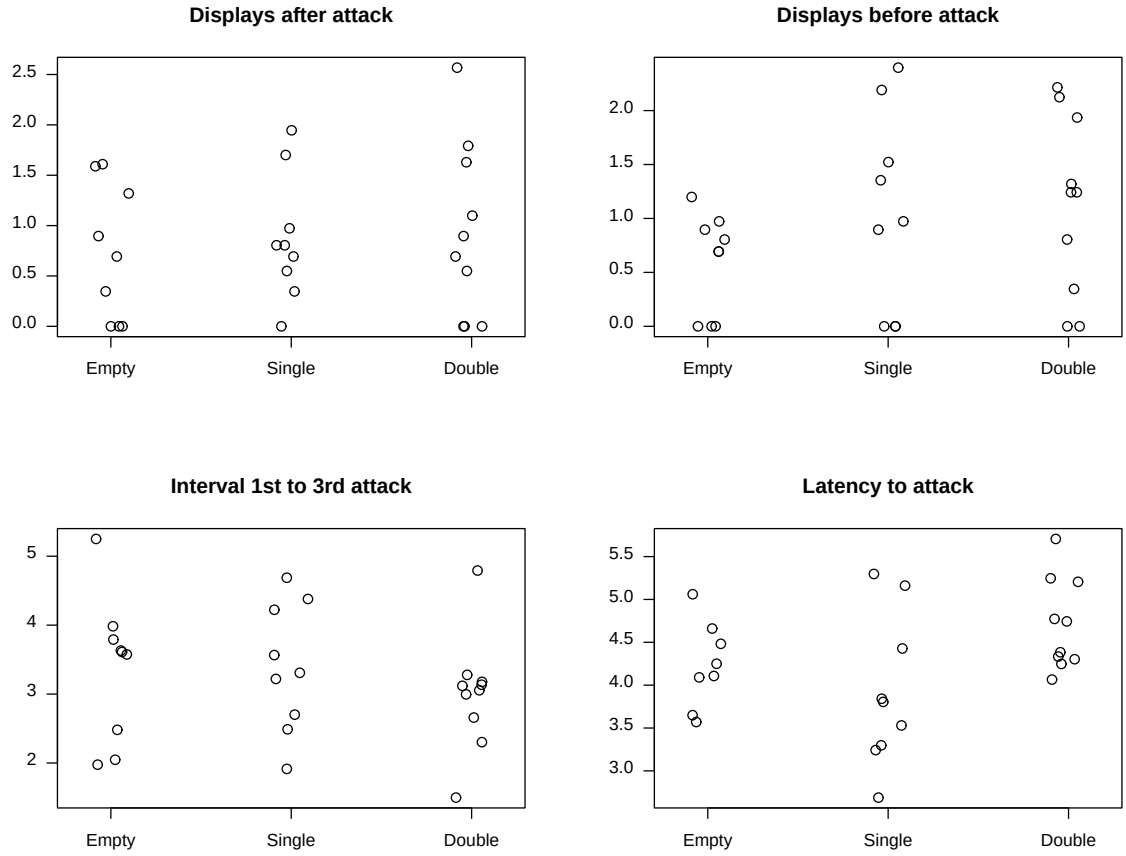

Figure 2.2: Aggressive behavior of resident male as a function of hormone treatment group. Explanation of variables in Table 2.2 (p. 30).
The y-axis is log(seconds) for the time variables and log(number +0.5) for number of displays before and after attack.

A MANOVA using the aggressive behavior variables (Table 2.2, p. 30) provided no evidence for effects of testosterone manipulations (Pillai's trace=0.37, $F_{8,42} = 1.20$, $p = 0.32$), as can be seen from Fig. 2.2 (p. 40). A multivariate regression using log testosterone and log corticosterone plasma levels and their interaction as independent variables provided no evidence of an interaction between plasma levels of testosterone and corticosterone (Pillai's trace=0.12, $F_{4,19} = 0.65$, $p = 0.64$) or a main effect of testosterone (Pillai's trace=0.24, $F_{4,20} = 1.59$, $p = 0.21$), and very weak evidence (Pillai's trace=0.29, $F_{4,20} = 2.06$, $p = 0.12$) that animals with higher plasma

levels of corticosterone are less aggressive in intruder encounters. In fact, test-wise p-values for the effects of corticosterone on aggressive behavior variables (Table 2.2, p. 30) are 0.0256 for latency to attack (log of latency to attack increases with log of corticosterone with slope ± s.e. 0.2806 ± 0.1164); p=0.0414 for interval between first and third attack (interval between first and third attack increases with increasing corticosterone (slope ± s.e: 0.4035 ± 0.1881); p=0.0641 for number of displays after attack (slope ± s.e. for corticosterone: -0.40 ± 0.20); p=0.7325 for number of displays before the first attack. Ordering the p-values and using Holm's method, the adjusted p-values for corticosterone are 0.1024, 0.1242, and 0.1282 (latency to attack, interval between first and third attack, number of displays after attack), suggesting that aggression might decrease with increasing corticosterone levels.

When examining the correlation between aggressive and antipredator behavior, a likelihood ratio test of the hypothesis that none of the canonical correlations was different from zero yields a p-value of 0.1236 ($\chi^2_8 = 12.67$ for association among aggressive and antipredator responses which, if anything, based upon the loadings, would suggest that animals that minimize risks from a predator are also those with higher aggressiveness). Thus, there is no evidence of an association between aggressive and antipredator behaviors.

There was no evidence of changes in mass or SVL in the experimental males (paired t-tests; comparison final and initial mass: $t_{27} = 0.31,\ p = 0.7579$; comparison final and initial SVL: $t_{27} = 1.04,\ p = 0.3085$). More importantly, there was no evidence that changes in mass or SVL were associated with hormone treatment group, plasma levels of testosterone, or plasma levels of corticosterone (all p-values$>$0.25).

2.4.2 Antipredator behavior: effects of hormonal manipulations and territorial intrusions

2.4.2.1 Approach and minimum distance

The model for log minimum distance provides strong evidence for period effects ($F_{1,80.3} = 14.66$, $p = 0.0003$) and territorial intrusion effects ($F_{1,80.3} = 14.62$, $p = 0.0003$); the model for approach distance provides strong evidence of period effects ($F_{1,65.4} = 9.73$, $p = 0.0027$), and of an interaction between territorial intrusion and approach speed ($F_{1,69.8} = 5.06$, $p = 0.0276$). In both cases, there is a decrease in distance with period, which suggests habituation: lizards allowed the predator to approach closer in later days of testing. For minimum distance, a territorial intrusion decreased minimum distance: lizards allowed a predator to approach closer before fleeing if the predator attacked 5 min after a territorial intrusion. For approach distance, if the predator attacked after a territorial intrusion lizards hid sooner when the predator's approach was faster (from a reparameterized model, regression coefficients ($\pm$ s.e.) for control and territorial intruder are 1.63 ($\pm$2.05) and 6.78 ($\pm$1.63); $t_{79.6} = 0.79$; $t_{78} = 4.17$; $p = 0.4315$ and $p < 0.0001$ respectively).

In addition, for both log approach and minimum distance there was evidence of an interaction between sequence and hormone treatment ($F_{2,21.6} = 4.31$, $p = 0.0266$, and $F_{2,22.1} = 4.74$, $p = 0.0194$) for approach and minimum distance respectively. As can be seen in Figure 2.3 (p. 44) both empty-implanted and single-implanted animals have smaller minimum distances when in sequence 1 (ICCI), whereas the pattern is reversed for doubly implanted animals; analogous results hold for approach distance.

The interpretation of a sequence by hormone interaction is complicated. First, the data show strong evidence of a hormone*sequence interaction but not of hormone*carry-over, hormone*territorial intrusion or hormone*period interactions. Differences among sequences can be the result of bad luck in the randomization process or of high-order carry-over effects (see also

Díaz-Uriarte 2000 b): a sequence term reflects all that is different among sequences that is not accounted for by treatment or carry-over effects. As is common in cross-over trials, we have used a very restrictive model for carry-over effects, which makes, among others, the assumption that there are only first-order carry-over effects. However, in this case it seems that the first period has an effect that lasts for the rest of the experiment.

To understand the results, we can analyze each period on its own for evidence of sequence*hormone interactions, which would be equivalent to territorial intrusion*hormone interactions, as within each period a sequence fully determines the type of territorial intrusion treatment. In these analyses, in periods one to three there was evidence (all p-values < 0.05) of a hormone by sequence (or territorial intrusion) interaction (in the fourth period the evidence is weak —$p = 0.13$). The test from the first period provided evidence of an interaction between territorial intrusion and hormone treatment; however, analyses of periods 2 to 4 confound the possible effect of a true territorial intrusion*hormone interaction with the effects of past events (first or higher-order carry-over effects). We can also examine if the difference (in the response variable) between the first and second period, between the second and third, between the third and fourth, and between the first and the mean of the other three, shows any evidence of sequence*hormone interactions; in other words, we can examine if the change in response variable from one period to the next is different among different hormone treatments. There was no evidence of interaction ($p > 0.09$ in all eight cases) or of main effects ($p > 0.13$ in all eight cases) of hormone treatment: after the first period, the change in response variable between one period and the following was not affected by hormone treatment. In other words, the change between periods is the same among levels of hormonal treatment, which means that effects of territorial intrusions are additive after the first period. Therefore, the interaction between sequence and hormone detected in the full model is the consequence of an interaction between territorial intrusion*hormone in the first period that is maintained for the rest of the study.

These patterns can be seen from Fig. 2.3 (p. 44) b & c: the trajectories over time are

roughly parallel across the three hormone treatment groups within each sequence. Within each sequence we can observe an overall decrease in response variable over time and a decrease in response variable corresponding to an intruder treatment. The parallel lines over time show the additive effects of period and intruder treatment. However, there were large differences among hormone treatment levels in the response in the first period: for both empty implanted and single implanted animals minimum (and approach) distance were smaller following a conspecific encounter, but in the doubly-implanted animals this pattern was reversed; if we compare only between empty implanted and single implanted the patterns were the same in both sequences, with the single implanted having larger approach and minimum distances than the empty implanted. In summary, in the first period the effect of a territorial intrusion depends on hormone treatment:, but after the first period effects of territorial intrusions and period act additively with respect to the value from the first period.

Given that the double-implanted group showed a behavior clearly distinct from the other two groups, and showed little overlap in their testosterone levels with the other two experimental groups, we next analyzed the data excluding the double-implanted animals. There was no

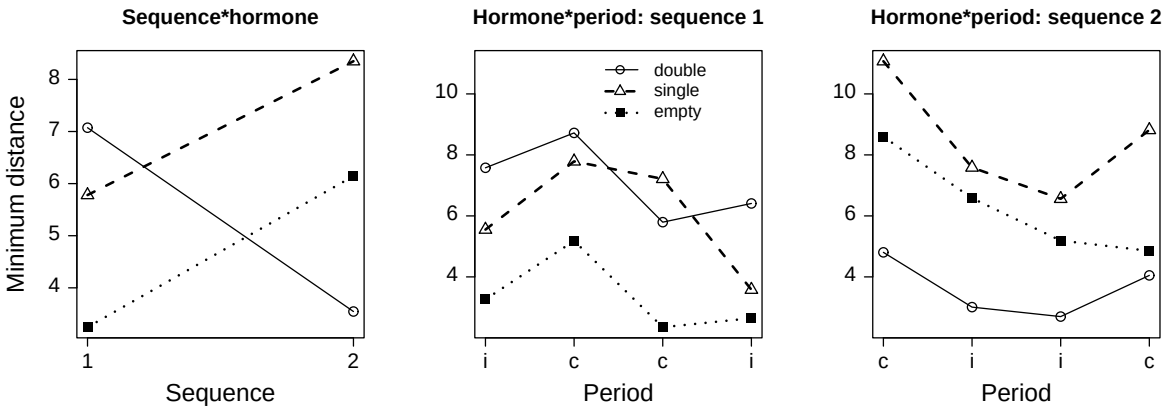

Figure 2.3: Minimum distance: effects of sequence, hormone, territorial intrusion and period. Each point corresponds to the (back-transformed) means from all the data available for the specified combination of factors. a) Interaction plot of sequence*hormone. b) and c) means by period for each hormone treatment and sequence group.

evidence of a hormone by sequence interaction ($F_{1,14} = 1.37$, $p = 0.2613$, and $F_{1,14} = 0.43$, $p = 0.5220$, for approach and minimum distance respectively). There was, however, evidence of differences between hormone treatments ($F_{1,15} = 6.74$, $p = 0.0203$, $F_{1,14} = 4.23$, $p = 0.0570$, for approach and minimum distance, respectively), where animals with a single implant have larger approach and minimum distances than the empty implanted males. The rest of the conclusions for effects of period, sequence, and territorial intrusion —or territorial intrusion by approach speed– remained unchanged (for approach distance p-values are 0.0480, 0.0042 and 0.0028 for sequence, period and territorial intrusion by approach speed; for minimum distance p-values are 0.0394, 0.0004 and 0.0014 for sequence, period and territorial intrusion).

Even though there was an effect of hormone treatment level on approach and minimum distance, when we fitted models that included log testosterone and corticosterone plasma levels instead of hormone treatment group for empty and single implanted animals we did not find any differences in either approach or minimum distance related to testosterone plasma levels ($p > 0.8$ for both minimum and approach distance). For approach distance, however, there was a significant interaction between corticosterone plasma levels and territorial intrusion ($F_{1,40} = 4.85$, $p = 0.0335$) where there was an increase in approach distance with increasing levels of corticosterone when animals were subject to a control presentation, but there was no change in approach distance with corticosterone plasma levels when animals were subject to a territorial intrusion (and, for an animal with a plasma corticosterone level equal to the observed mean corticosterone plasma level, approach distance is smaller when exposed to an intruder than when exposed to a control presentation). There was no evidence of such an interaction for minimum distance ($F_{1,50} = 0.81$, $p = 0.37$). None of these conclusions are changed by applying Holm's multiple comparisons approach.

2.4.2.2 Time to reemerge and time to full exposure

Analyses of time to reemerge using hazard models provided strong evidence of differences among hormone level treatment groups ($\chi^2_2 = 14.61$, $p = 0.0007$) where double-implanted animals reemerged later than the empty implanted and single implanted reemerged slightly sooner than the empty implanted. Analyses of time to full exposure yielded results in the same direction, although not significant ($\chi^2_2 = 4.74$, $p = 0.0934$). If we specifically test for differences between the double implanted and the other two groups, there was evidence of differences for both response variables ($\chi^2_1 = 3.79$, $p = 0.052$,$\chi^2_1 = 7.21$, $p = 0.0072$, for time to reemerge and time to full exposure, respectively). If we exclude the double-implanted animals, there was no evidence of differences between empty and single implanted for any of the two response variables ($\chi^2_1 = 1.74$, $p = 0.19$, $\chi^2_1 = 0.46$, $p = 0.50$, for time to reemerge and time to full exposure, respectively; the hazard rate for a double implanted animal was 0.63 that of any of the other two groups for time to reemerge, and 0.392 for time to full exposure). Analyses using testosterone and corticosterone plasma levels from empty and single-implanted animals did not show any evidence of differences in either response variable related to hormone plasma levels (for testosterone both p-values > 0.6; for corticosterone both p-values > 0.19).

For both response variables, there was very strong evidence that being subject to a territorial intrusion resulted in faster reemergence ($\chi^2_1 = 18.19$, $p < 0.0001$, $\chi^2_1 = 15.92$, $p < 0.0001$, for time to reemerge and time to full exposure, respectively, in analyses that include the three hormone treatment groups ($\chi^2_1 = 22.9$, $p < 0.0001$, $\chi^2_1 = 11.61$, $p < 0.0011$, for time to reemerge and time to full exposure, respectively, in analyses that include only empty and single implanted animals), as shown in Fig. 2.4 (p. 47). The hazard rate of an animal exposed to a conspecific intrusion is 3.78 times that of an animal following a control presentation for time to reemerge, and 2.71 for time to full exposure (from analyses that exclude the double implanted animals; similar results are obtained from analyses with all three groups). None of these conclusions are changed by applying Holm's multiple comparisons approach.

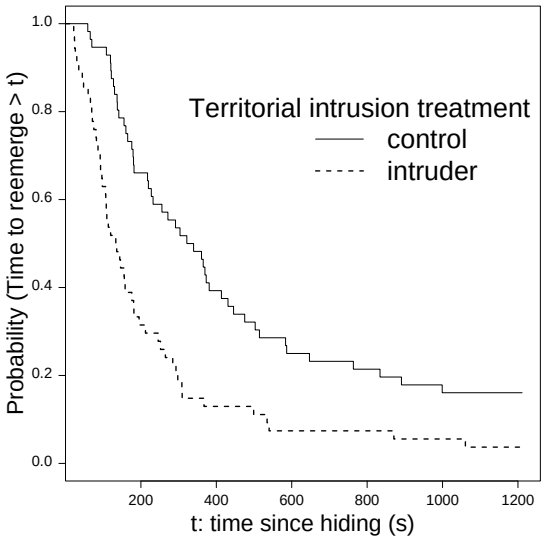

Figure 2.4: Survival curves of time to reemerge based on the Kaplan-Meier estimator of the survival function. The y-axis can also be understood as "probability of not having reemerged by time t".

2.5 Discussion

The initial hypothesis that testosterone would modify antipredator behavior, so that animals with higher testosterone levels would incur greater predation risks, was based upon the assumption that testosterone results in increased allocation to territorial defense (see Introduction). In this study, the lack of an effect of testosterone treatment on aggressive behavior suggests, to the contrary, that in male *T. hispidus* testosterone does not result in an increase in the allocation to territorial defense (at the expense of other costs). Thus, there is no reason to expect an increase in exposure to predation related to testosterone manipulations in males of this species, or an interaction between testosterone treatment and the effects of a past territorial intrusion. Our results show, for empty and single implanted animals, that increased testosterone was not associated with increased exposure to predation; in fact, increased testosterone resulted in decreased exposure to predation as measured by changes in approach and minimum distance. Animals with double implants exhibited a different pattern; first, they were the ones that took

significantly longer to reemerge; second, and more strikingly, the effect of sequence (type of first trial —control or intruder) on approach distance was opposite to that observed in the other two groups (see Fig. 2.3, p. 44). These differences between the double implanted animals and the other two groups could be related to pharmacological effects of the double testosterone implant (the double implanted animals are outside the range of testosterone levels for wild animals; see Fig. 2.1, p. 39) and also to the effects of the testosterone implants on other hormones, such as corticosterone; together, these changes might affect the response to stimulae in ways that differ from the other two treatment groups.

The above results do not preclude changes in antipredator behavior, and interactions between testosterone levels and effects of past territorial intrusions on antipredator behavior, in the direction predicted in the introduction where there is an increase in aggression with testosterone (e.g., *Sceloporus jarrovi*: Moore & Marler, 1987; *Anolis sagrei*: Tokarz, 1987, 1995; *Uta stansburiana*: DeNardo & Sinervo, 1994). Our results suggest that testosterone does not play a role in the aggressive behavior of male *T. hispidus*, a tropical lizard with flexible breeding patterns (from marked seasonality —e.g., Prieto et al., 1970— to extended breeding seasons —Vitt & Goldberg, 1983; pers. obs.), where both male and female are territorial year around, at least in the study area. We are not aware of other studies on the effects of testosterone manipulations in hormone levels of tropical lizards, but studies with tropical birds that are territorial year around have yielded mixed results (see Hau et al., 2000; Wikelski et al., 1999; and references therein) indicating that testosterone might not necessarily play a role in the aggressive behavior of tropical vertebrates that are aggressive throughout the year. However, a role of testosterone on the aggressive and territorial behavior cannot be excluded without additional studies involving castration (e.g., Moore & Marler, 1987), and/or antiandrogen treatment (e.g., Tokarz, 1987). Moreover, the lack of effects of testosterone manipulations on aggressive behavior in this experiment could be related to the already elevated testosterone plasma levels of the empty implanted animals compared to the wild animals (e.g., Fig. 2.1, p. 39). The differences between wild males and empty implanted males can be caused by the housing conditions, in

particular the close proximity of a female during three weeks, a regular food supply, and not being challenged by other males for three weeks.

Our work, though, suggest that corticosterone could play a role on how past aggressive interactions affect antipredator behavior. First, there was some evidence that increased in corticosterone resulted in decreased aggression towards intruders; second, the interaction between corticosterone and territorial intrusion on approach distance indicates that, in the absence of past conspecific intrusions, increased corticosterone is associated with decreased exposure to predation, but that these effects of corticosterone can be overridden by a past territorial intrusion (since corticosterone was not associated with approach distance following a conspecific intrusion).

The results of this study provide additional confirmation (on the same species) of the results in Díaz-Uriarte (1999): animals increased their exposure to predators following an aggressive encounter. However, in this study we found this effect in both approach distance and reemergence behavior, whereas Díaz-Uriarte (1999) only found this effect on reemergence behavior. Recent theoretical work (Díaz-Uriarte, 2000c) indicates that the increased cost of hiding following a conspecific intrusion should only modify reemergence behavior, not when to hide from an attacking predator. However, other factors can operate together with increased cost of hiding (estimation of the probability that the approaching predator is an attacking one, interrupted foraging and environmental sampling) that result in a decrease in approach distance. Most of the experimental conditions of both studies were very similar, but three differences could explain lack of detection of effects on approach distance in Díaz-Uriarte (1999): (1) smaller sample size (15 vs. 28 animals); (2) smaller number of measures per individual (two vs. four); (3) faster approach speed 0.42 m/s vs. 0.22 m/s). Because of the first two differences this experiment had higher statistical power than the one in Díaz-Uriarte (1999); slower approach speed in this experiment means that the same decrease in approach distance does not result in the same increase in mortality risk, and therefore other factors (e.g., estimation of the probability

of an attack by the predator) could have detectable effects in hiding.

These differences in experimental conditions might also explain why we found sequence effects in approach and minimum distance in this experiment, but none were found in Díaz-Uriarte (1999). A more likely explanation, though, is the difference in the time that experimental males are isolated from other conspecific males before the tests were conducted. In Díaz-Uriarte (1999) animals were precluded from fighting with other conspecific males for one week. In contrast, in this experiment males were isolated for about three weeks, and thus an aggressive interaction could have a much larger and longer lasting effect on antipredator behavior, and explain why animals from the ICCI (intruder/control/control/intruder) sequence showed, overall, smaller approach and minimum distances. These long lasting effects, however, were not affected by testosterone manipulations. These long lasting effects could not have been detected with other types of designs, and should be taken into account in future studies.

In spite of the strong effects of a past aggressive interaction on antipredator behavior, we found no covariation between aggressive and antipredator behaviors (i.e., males that showed more aggressive behavior towards conspecifics did not show bolder behavior towards a predator). These data, thus, constitute a counter-example of the idea that a correlation between aggressive and antipredator behavior could share a common physiological basis and be widespread in nature (Reichert & Hedrick, 1993). Male *T. hispidus* cannot be positioned along a single shy-bold axis, where aggression and antipredator responses are essentially the manifestation of an underlying "fearfulness" trait (Huntingford, 1976; see also Wilson et al., 1993, 1994). The lack of correlation between aggressive and antipredator behavior is not inconsistent with the strong effects of a past aggressive interaction on antipredator behavior. The latter are based on within-individual effects, whereas the former relate to among-individual covariation in aggressive and antipredator responses. Moreover, the lack of among-individual covariation in antipredator and aggressive behavior does not exclude that, within individuals, an increase in the aggressiveness of the interaction could result in a larger increase in predator exposure.

2.6 Acknowledgments

We thank C. Lázaro-Perea for discussion and help building, cleaning, and taking the enclosures down, assistance during surgery that allowed this study to be blind, and comments on the ms.; Patricia Martin for performing the radioimmunoassays; Departamento de Fisiologia, Universidade Federal do Rio Grande do Norte, Brazil, and in particular M. F. Arruda and M. B. C. de Souza for invaluable logistic support; Eflex-IBAMA of Nisia Floresta, specially J. Dantas, for facilities, tools, and permission to perform this work; the owners of the Facenda Luis Carlos (Santo Antonio, RN) for permission to work there; E. Nascimento and several children of Porto and Nisia Floresta for assistance with lizard capture; the Wisconsin Regional Primate Research Center's Assay Lab for facilities and technical assistance.

2.7 References

Abrahams, M. V., and Pratt, T. C. 2000. Hormonal manipulations of growth rate and its influence on predator avoidance-foraging trade-offs. Canadian Journal of Zoology, 78, 121-127.

Abrahams, M. V., and Sutterlin, A. 1999. The foraging and antipredator behaviour of growth-enhanced transgenic atlantic salmon. Animal Behaviour, 58, 933,942.

Bouissou, M. F., and Vandenheede, M. 1996. Long-term effects of androgen treatment on fear reaction in ewes. Horm. Behav. 30:93-99.

Cawthorn, J. M., Morris, D. L., Ketterson, E. D., and Nolan, V. 1998. Influence of experimentally elevated testosterone on nest defence in dark-eyed juncos. Animal Behaviour, 56, 617-621.

Chandler, C. R. 1995. Practical considerations in the use of simultaneous inference for multiple tests. Animal Behaviour, 49, 524-527.

Collett, D. 1994. Modelling Survival Data in Medical Research. London: Chapman & Hall.

DeNardo, D. F., and Sinervo, B. 1994. Effects of steroid hormone interaction on activity and home-range size of male lizards. Horm. Behav. 28:273-287.

Díaz-Uriarte, R. 1999. Anti-predator behaviour changes following an aggressive encounter in the lizard *Tropidurus hispidus*. Proceedings Royal Society of London, Series B, 266, 2457-2464.

Díaz-Uriarte, R. 2000a. Cross-over trials in animal behaviour. I: Misuse, carry-over effects, and design. In review, Animal Behaviour

Díaz-Uriarte, R. 2000b. Cross-over trials in animal behaviour. II: Analysis and plotting. In review, Animal Behaviour

Díaz-Uriarte, R. 2000c. Territorial intrusion risk and antipredator behaviour: a mathematical model. In review, Proceedings Royal Society of London, Series B.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. 1994. Analysis of Longitudinal Data. Oxford: Oxford University Press.

Foufopoulos, J., Díaz-Uriarte, R., and Marler, C. A. 2000. Testosterone effects on lizard parasite loads as revealed through experimental manipulations. Behavioral Ecology, in review.

Hau, M., Wikelski, M., Soma, K. K., and Wingfield, J. C. 2000. Testosterone and year-round territorial aggression in a tropical bird. General and Comparative Endocrinology, 117, 20-33.

Huntingford, F. A. 1976. The relationship between anti-predator behaviour and aggression among conspecifics in the three-spined stickleback, Gasterosteus aculeatus. Anim. Behav. 24:245-260.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monog. 54:187-211.

Johnsson, J. I., Petersson, E., Jonsson, E., and Bjornsson, B. T. 1999. Growth hormone-induced effects on mortality, energy status and growth: a field study on Brown trout (*Salmo trutta*). Functional Ecology, 13, 514-522.

Johnsson, J. I., Petersson, E., Jonsson, E., Bjornsson, B. T., and Jarvi, T. 1996. Domestication and growth hormone alter antipredator behaviour and growth patterns in juvenile brown trout, *Salmo trutta*. Canadian Journal of Fisheries and Aquatic Sciences, 53, 1546-1554.

Jones, B., and Kenward, M. G. 1989. Design and analysis of cross- over trials. Chapman & Hall, New York.

Klein, J. P. and Moeschberger, M. L. 1997. Survival Analysis. New York: Springer-Verlag.

Krzanowski, W. J. 1990. Principles of Multivariate Analysis. New York: Oxford University Press.

Lee, E. W., Wei, L. J., and Amato, D. A. 1992. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Survival analysis: state of the art. (Ed. by J. P. Klein and P. K. Goel), pp. 237-247. The Netherlands: Kluwer Academic

Lin, D. Y. 1994. Cox regression analysis of multivariate failure time data: the marginal approach. Statistics in Medicine, 13, 2233-2247.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. 1996. SAS System for Mixed Models. Cary, NC: SAS Institute.

Marler, C. A., and Moore, C. M. 1988. Evolutionary costs of aggresion revealed by testosterone manipulations in free-living male lizards. Behav. Ecol. Sociobiol. 23:21-26.

Marler, C. A., and Moore, M. C. 1991. Supplementary feeding compensates for testosterone-induced costs of aggression in male mountain spiny lizards, Sceloporus jarrovi. Anim. Behav. 42:209- 219.

Moore, M. C. 1988. Testosterone control of territorial behavior: tonic-release implants fully restore seasonal and short-term aggressive responses in free-living castrated lizards. Gen. Comp. Endocrinol. 70:450-459.

Moore, M. C., and Marler, C. A. 1987. Effects of testosterone manipulations on nonbreeding season territorial aggression in free-living male lizards, Sceloporus jarrovi. Gen. Comp. Endocrinol. 65:225-232.

Morrison, D. F. 1990. Multivariate Statistical Methods, 3r Ed. New York: McGraw-Hill.

Pinheiro, J. C & Bates, D. M. 2000. Mixed-Effects Models in S. Springer-Verlag, in press.

Prieto, A. S., León, J. R., and Lara, O. 1976. Reproduction in the tropical lizard, *Tropidurus hispidus* (Sauria: Iguanidae). Herpetologica 32:318-323.

Reichert, S. E., and Hedrick, A. V. 1993. A test for correlations among fitness-linked behavioural traits in the spider Agelenopsis aperta (Araneae, Agelenidae). Anim. Behav. 46:669-675.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution, 43, 223-225.

Rodrigues, M. T. 1987. Sistemática, ecologia e zoogeografia dos Tropidurus do grupo Torquatus ao sul do rio Amazonas (Sauria, Iguanidae). Arquivos de Zoologia, S. Paulo 31:105-230.

Salvador, A., Veiga, J. P., Martin, J., and Lopez, P. 1997. Testosterone supplementation in subordinate, small male lizards: consequences for aggressiveness, color development, and parasite load. Behav. Ecol. 8:135-139.

Sih, A. 1992. Integrative approaches to the study of predation: general thoughts and a case study on sunfish and salamander larvae. Ann. Zool. Fenn. 29:183-198.

Sinervo, B., and Basolo, A. L. 1996. Testing adaptation using phenotypic manipulations. Pp. 149-185 in Rose, M. R., and Lauder, G. V. eds., Adaptation, Academic Press, Boston.

Stamps, J. A. 1991. Why evolutionary issues are reviving interest in proximate behavioral mechanisms. Amer. Zool. 31:338-348.

Therneau, T. and Grambsch, P. 2000. Modeling Survival Data: Extending the Cox Model. Springer-Verlag, New York, in press.

Tokarz, R. R. 1987. Effects of the antiandrogen cyproterone acetate and flutamide on male reproductive behavior in a lizard (Anolis sagrei). Horm. Behav. 21:1-16.

Tokarz, R. R. 1995. Importance of androgens in male territorial acquisition in the lizard Anolis sagrei: an experimental test. Anim. Behav. 49:661-669.

Tulley, J. J., and Huntingford, F. A. 1988. Additional information on the relationship between intra-specific aggression and anti-predator behaviour in the three-spined stickleback, Gasterosteus aculeatus. Ethology 78:219-222.

Vandenheede, M. and Bouissou, M. F. 1996. Effects of castration on fear reactions of male sheep. Applied Animal Behaviour Science, 47, 211-224.

Van Sluys, M. 1993. The reproductive cycle of *Tropidurus itambere* (Sauria: Tropiduridae) in southeastern Brazil. J. Herpetol., 27, 28-32.

Verbeke, G. and Molenberghs, G (eds).1997. Linear mixed models in practice. A SAS-Oriented approach. New York: Springer-Verlag

Vitt, L. J. 1995. The ecology of tropical lizards in the caatinga of northeast Brazil. Occasional Papers of the Oklahoma Museum of Natural History 1:1-29.

Vitt, L. J., and Goldberg, S. R. 1983. Reproductive ecology of two tropical iguanid lizards Tropidurus torquatus and Platynotus semiateniatus. Copeia 1983:131-141.

Wei, L. J., Lin, D. Y., and Weissfeld, L. 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association, 84, 1065-1073.

Wikelski, M., Hau, M., and Wingfield, J. C. 1999. Scial instability increases plasma testosterone ina year-round territorial neotropical bird. Proceedings Royal Society of London, Series B, 266, 551-556.

Wilson, D. S., Clark, A. B., Coleman, K., and Dearstyne, T. 1994. Shyness and boldness in humans and other animals. TREE 9:442-446.

Wilson, D. S., Coleman, K., Clark, A. B., and Biederman, L. 1993. Shy-bold continuum in pumpkinseed sunfish (Lepomis gibbosus): an ecological study of a psychological trait. J. Comp. Psychol. 107: 250-260.

Wright, S. P. 1992. Adjusted P-values for simultaneous inference. Biometrics, 48, 1005-1013.

Chapter 3

Territorial intrusion risk and antipredator behaviour: a mathematical model

3.1 Abstract

In territorial animals that hide to avoid predators, a predatory attack creates a conflict because a hiding animal cannot defend its territory from conspecific intruders. When intruders are persistent, a past conspecific intrusion informs a territorial resident that future intrusions by the same animal are likely. Using a mathematical model, I examine the effects that past territorial intrusions can have on antipredator behaviour when intruders are persistent. Past aggressive intrusions rarely affect time to hide: the optimal behaviour is to hide as soon as the predator initiates its attack. Time to reemerge is strongly affected by past aggressive interactions (animals reemerge sooner from a refuge), and these effects depend on the time of the predator's attack, the reintruder's pattern of return, and the intrusion rates of other conspecifics. Differences between my findings and those from previous studies suggest that the trade-off between antipredator behaviour and territorial defence can involve different types of costs than the trade-off antipredator behaviour- foraging. The results of this model establish a connection between population level processes, mating system and defensibility of resources, and antipredator behaviour, and can have empirical and theoretical relevance for studies of the

(co)evolution and ecological consequences of aggressive and antipredator strategies.

3.2 Introduction

The antipredator strategy of territorial animals should be affected by the need to defend a territory. Theoretical and empirical work on the trade–off between predator avoidance and foraging has shown that antipredator behaviour will change when there are changes in the terms of the trade–off between mortality risk from predation and costs of hiding/escaping from predators (see Clark, 1994; Ydenberg & Dill, 1986; reviews in Lima & Dill, 1990; Lima, 1998). For instance, animals adopt behavioural strategies that lead to an increase in exposure to predation (e.g., delaying escape from a predator) when the costs of interrupting foraging increase (e.g., when foraging at a better patch).

In contrast to the wealth of studies on the trade-off between antipredator and foraging behaviour, there is little research on the trade–off between antipredator behaviour and territorial defence, even though the reproductive success of territorial animals can be strongly affected by successful territorial defence. The approach of a predator creates conflicting demands on a territorial animal: hiding minimises mortality from predation but decreases the chances of detecting and chasing away conspecific intruders (i.e., increases the territorial costs of hiding). There is evidence that increases in predation risk tend to result in a decrease in the number or intensity of aggressive interactions (e.g., Baker et al., 1999; Brick, 1999; Helfman, 1989; Krupa & Sih, 1998; Martel, 1996; Whitehouse, 1997; Wisenden & Sargent, 1997), but the effects of aggressive interactions on antipredator behaviour have been rarely examined (but see Brick, 1998; Cooper, 1999; Díaz-Uriarte, 1999; Jakobsson et al., 1995). The trade–off between territorial defence and predator avoidance can be particularly interesting if there are short–term changes in the territorial costs of hiding that are caused by local changes in the social environment. In fact, in some territorial species intruders enlarge or obtain territories by intruding persistently into the territories of settled animals (review in Stamps & Krishnan,

1995, 1998). In these cases a recent conspecific intrusion indicates an increased probability of future intrusions and therefore the territorial costs of hiding could be very high following a conspecific intrusion; thus, antipredator behaviour should change to decrease the chances of territorial intrusions at the cost of increased predation risks.

There is recent empirical evidence (Díaz-Uriarte, 1999; Díaz-Uriarte & Marler, in prep.) that territorial males of the lizard *Tropidurus hispidus* increase their exposure to predation when a predator approaches shortly after the territorial male has chased away a conspecific intruder male, consistent with the arguments above. In these experiments, male lizards were presented (and allowed to fight) with a conspecific intruder male, and 5 min later were subject to a simulated predatory attack by a human. Antipredator behaviour was characterised using two types of variables: 1) when did the lizard initiate escape from the predator; 2) when did the lizard reemerge from the refuge after hiding. In the first study, only time to reemerge from a refuge is affected by past aggressive interactions; in contrast, initiation of hiding does not depend on past aggressive interactions. In the second study, both time to reemerge from a refuge and initiation of hiding are affected by past aggressive interactions.

The conditions that give rise to a trade-off between antipredator and territorial behaviour in males of the lizard *Tropidurus hispidus* are likely to be common to many other species that are both territorial and prey of other animals. Thus effects of past aggressive interactions on antipredator behaviour are likely to be widespread, but demographic and social factors that vary both within and among species, such as population density and behaviour of reintruders, should affect this trade-off. The purpose of this paper is to investigate how past aggressive interactions should affect antipredator behaviour in territorial animals that need to defend their territories against conspecifics and are also potential prey that use refuges to avoid predation. The model focuses on the effects of the reintruder's behaviour, the probability of intrusion of other conspecifics, and the timing of predator attack relative to the end of the conspecific intrusion.

3.3 The model

3.3.1 The basic problem

Suppose that a territorial male is defending an area that overlaps the home ranges of several females. If other males invade the territory while the resident is hiding they could mate with the females in the territory and the number of females that can be fertilised by the invading males increases with the time these invading males spend in the territory before being evicted. This male chases away a conspecific intruder at time 0. Some time later (t_p) a predator initiates approach (the predator is detected as soon as it initiates approach). The resident needs to decide: (1) when to escape (time to hide, t_h), and (2) when to reemerge (time to reemerge, t_r). The longer the resident waits to hide (i.e., the larger the t_h) or the shorter the time to reemerge, the more likely it is to be killed by the predator. On the other hand, the longer the animal remains hiding the more likely it is that intruders can invade the territory. Once an intruder enters the territory, it stays there until the resident reemerges, and the reproductive success of the resident decreases with time that intruders spend in its territory. There are two types of intruders, the re-intruder that was chased away at time 0 and other conspecifics from the overall population. The effects of the prior aggressive encounter (the animal chased away at time 0) are only related to the probability that the reintruder returns, but do not affect the rate of intrusion of other conspecifics. Intruders cannot successfully invade the territory if the predator is in the area or if the territorial resident is not hiding, but they can attempt to reinvade during these periods. The lack of attempted reinvasion by the reintruder prior to the resident hiding can provide the resident with information on the probability of a reinvasion in the future.

I assume that the resident has to maximise fitness, the product of its probability of surviving the attack of the predator times its expected reproductive success, by choosing optimal values of time to hide (t_h) and time to reemerge (t_r). In the next sections I give details about each

component of the model (see also Table 3.1, p. 62, for summary of variables). In this model, I make many simplifying assumptions, with function selection dictated by the desire to have simple functions that are, nonetheless, biologically plausible.

3.3.2 Surviving the predator's attack

The main biological assumptions that I make with respect to the predator's attack are: (1) that survivorship is a monotonically decreasing function of time to hide; (2) that survivorship is a monotonically increasing function of time to reemerge; (3) that the predator's attack is a fast event; (4) that the decrease in survivorship from delaying hiding for one unit of time is larger than the decrease in survivorship from reemerging one unit earlier for sufficiently large values of time to reemerge. I have implemented these as follows.

The probability of surviving the initial attack of the predator decreases linearly with t_h so that at $t_h \geq 10$ the probability of surviving is 0. Thus, the probability of surviving the initial attack is

$$1 - \frac{t_h}{10} \tag{3.1}$$

for all $0 \leq t_h \leq 10$, and 0 otherwise. Once the resident hides in the refuge, the predator stays around the area but has a constant rate of leaving ρ (thus, the predator's time of leaving is an exponential distribution with mean $1/\rho$). I assume that the resident is killed if it reemerges from the refuge while the predator is in the area. Thus, the probability that the resident survives reemergence is the probability that the predator has left the area by t_r or

$$1 - \mathrm{e}^{-\rho t_r}. \tag{3.2}$$

The probability of surviving the attack is therefore the product of expressions 3.1 and 3.2. There is no mortality while the resident is hiding.

3.3.3 Time that intruders spend in the territory

I assume that the decrease in reproductive success of the resident is a linear function of the time that intruders spend in its territory. Final reproductive success is

$$I - c\,\text{Total time intruders spend in territory} \tag{3.3}$$

where I is the initial value or initial territorial assets (i.e., the reproductive success yielded by a territory before any intruder spends any time at all, or before any intruder causes any decrease) and c is a scaling factor for the rate of decrease of reproductive success with time that intruders spend in the territory (the larger c the greater the decrease in reproductive success per unit time that intruders spend in the territory).

3.3.3.1 Time spent by other conspecifics

I assume that the only variable that affects reproductive success is the total accumulated time that intruders spend in the territory (i.e., one intruder spending 20 time units in the territory results in the same decrease in reproductive success as four intruders each spending 5 time units). I model the entry of the other conspecifics (as opposed to the re-intruder) as a Poisson process, where β is the rate of entry of intruders, and does not change over time or with the number of intruders already in the territory (except that no conspecific can intrude in the territory if the predator is still present). It is shown in Appendix 3.A (p. 78) that the expected total time that the other conspecifics accumulate is given by

$$\int_0^{t_r} \beta \frac{(t_r - s)^2}{2} \rho\, \mathrm{e}^{-\rho s}\, \mathrm{d}s = \frac{\beta}{\rho^2} - \frac{\beta}{\rho^2 \mathrm{e}^{\rho t_r}} - \frac{\beta t_r}{\rho} + \frac{\beta t_r^2}{2}. \tag{3.4}$$

3.3.3.2 Time spent by the re-intruder

In contrast to the other conspecifics, the re-intruder is the one particular individual that was chased away at time 0. The reintruder can either attempt to reinvade the territory or not; if it attempts a reinvasion, the reintruder's attempted return time has a certain probability density function (pdf). However, the reintruder can only reinvade successfully if the resident is hiding and the predator has left the area. I show in Appendix 3.A (p. 78) that if the time of return t_i is distributed according to the pdf $f_T(t_i)$ (and $F_T(t_i)$ is the cumulative distribution function) the expected time that the reintruder spends in the territory (x) is given by

$$E[X|\text{No attempted invasion by } t_p + t_h] = \frac{p}{1 - pF_T(t_p + t_h)} \int_0^{t_r} x f_T(t_p + t_h + t_r - x)\,(1 - \mathrm{e}^{-\rho(t_r - x)})\,\mathrm{d}x. \tag{3.5}$$

I evaluated this integral numerically.

3.3.4 Parameter values and robustness of results

The range of values for the different parameters is shown in Table 3.1 (p. 62). Changes in the values of one or more parameters only lead to numerical differences, but not to changes in qualitative patterns (see also Discussion); for instance, notice how the different panels within Fig. 3.2, 3.3, 3.4 (pp. 65, 66, 67) are scaled versions of each other. The only exception to this are very small values for the variance of reintruder's return time (see below). All results shown in the figures correspond to a probability of reintrusion (p) of 0.9; changes in this parameter only either increase or decrease the effects of the reintruder, but in most of the cases examined effects of a past reintrusion are observable with $p = 0.4$.

For fixed reintrusion time I have arbitrarily set $t_i = 400$. Choosing a different value makes no difference, as the relevant variable is not t_i (or its mean for the log-normal distribution), but $t_{ip} (= t_i - t_p)$, the time at which the reintruder returns with respect to the predator attack. To

Table 3.1: Main variables and parameters of the model.

Symbol	Meaning	Range
t_h	Time to hide (relative to initiation of predator attack)	Optimised variable
t_r	Time to reemerge (relative to initiation of hiding)	Optimised variable
t_p	Time to predator attack (relative to time when intruder is chased away)	0-7000
p	Probability of reintruder's return	0.4-0.99
t_i	Time of reintruder's return (relative to time when intruder is chased away)	400 or random
μ	For reintruder with log-normal return pdf: mean of log(return time)	Log(400)
σ	For reintruder with log-normal return pdf: standard deviation of log(return time)	0.001-1
λ	For reintruder with exponential return pdf: mean of the exponential distribution	2-800
β	Rate of intrusion of other conspecifics	0.00009-0.012
ρ	Rate of predator leaving the area after resident hides; mean time to leave $= 1/\rho$.	0.005-0.05
I	Initial territorial assets	0.1-4
c	Rate of decrease of reproductive success with time intruders spend in territory	0.02-0.9
	Variables derived from the above	
t_{ip}	Time of reintruder's return relative to time of predator's attack ($t_{ip} = t_i - t_p$)	
h	Time of hiding ($h = t_p + t_h$)	
r	Time of reemergence ($r = t_p + t_h + t_r$)	

examine the effects of variation in the intruder's behaviour, I have modelled return times using two different distributions, an exponential and a log-normal, and have generated additional variability in the reintruder's behaviour by modifying the parameters of these pdf's. With the log-normal pdf, the first parameter (= the mean of the log (return time)) has been set equal to log(400), to make it comparable to the fixed reintrusion time case, and I have varied the second

parameter, the standard deviation of log(return time) (note that the mean of t_i is not exactly 400). For the exponential, I have changed its mean, which also changes its variance (since for an exponential distribution the variance is the square of the mean). Several examples of the pdf of return times are shown in Fig. 3.1.

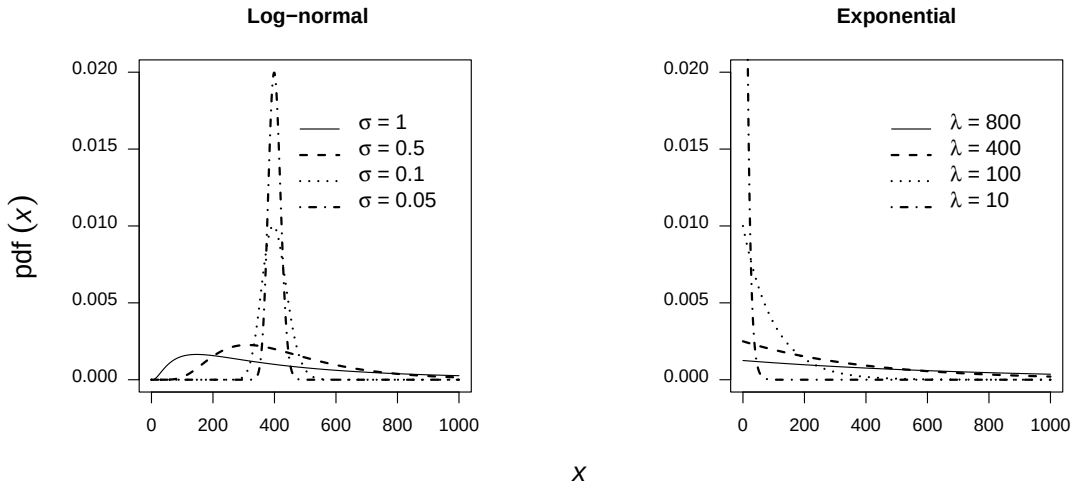

Figure 3.1: Examples of the probability density functions (pdf's) used for the reintruder's return time.

3.4 Results

The focus of this work is the effect of a past reintrusion, which can be evaluated comparing the optimal values of t_h and t_r with the optimal values for an identical situation without reintruder (i.e., when only other conspecifics can invade). Thus, I first examine the effects of having only other conspecifics on optimal t_h and t_r. Next, I show the results when a reintruder is added. Since the most relevant results are those from a reintruder with stochastic behaviour, I concentrate on those; the results for a reintruder with fixed return time are shown in Appendix 3.B (p. 82).

3.4.1 Effects of other conspecifics

When there are no re-intruders, but only other conspecifics, nothing is gained by delaying hiding from an attacking predator. For fixed β (rate of intrusion of other conspecifics) and ρ (predator's rate of leaving the area), the only variable that determines the time that intruders spend in the territory is time to reemerge (t_r; see expression 3.4, p. 60), and delaying hiding ($t_h > 0$) only results in increased mortality risk. Given a fixed loss in reproductive success caused by other conspecifics (i.e., for a fixed t_r), this loss can be kept constant keeping t_r constant, but survivorship maximised by hiding at 0. Thus, we cannot find, for any t_r, any $t_h > 0$ that will be better than $t_h = 0$, and hence the optimal option is to always hide at $t_h = 0$. (For the other conspecifics no information can be gained by delayed hiding, since the probability of invasion of other conspecifics is independent of past events; this differs from the situation with a reintruder, where information can be gained about the probability of a future return —see below). In contrast to time to hide, other conspecifics do influence time to reemerge. Increases in β and ρ decrease t_r^* (optimal time to reemerge): if the rate of intrusion is higher the resident ought to reemerge sooner (at the expense of survivorship), and if the predator is likely to leave the area sooner, the resident can reemerge sooner without incurring increased predation risks. If intruders have a large depressing effect on reproductive success (large c —compare a vs. b and c vs. d in Fig. 3.2, p. 65) or if initial assets (I) are small —compare a vs. c and b vs. d in Fig. 3.2—, the resident will reemerge sooner.

3.4.2 Effects of the reintruder

We now add a reintruder and examine how optimal time to hide and optimal time to reemerge change relative to the optimal time to hide and time to reemerge when there are only other conspecifics (previous section).

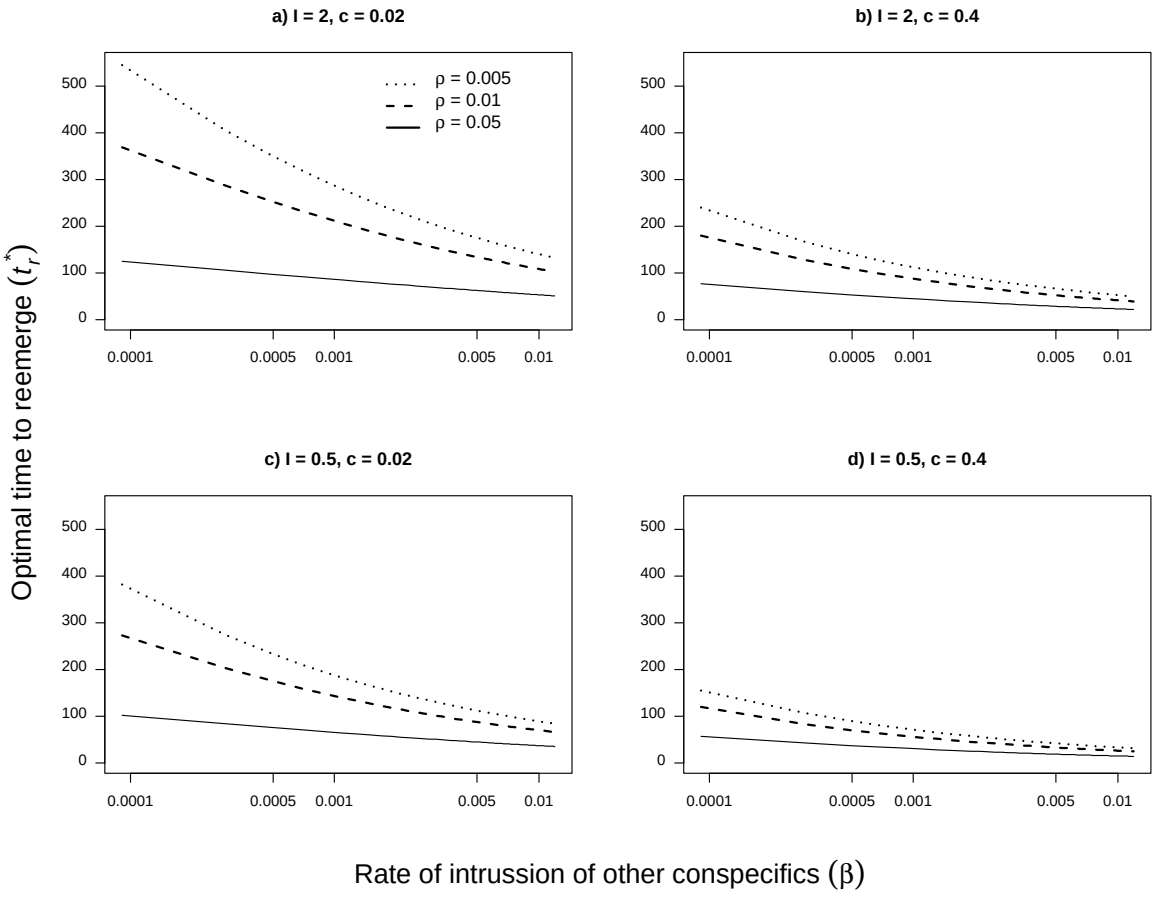

Figure 3.2: Optimal time to reemerge (t_r) when there is no reintruder, as a function of rate of intrusion of other conspecifics (β), for different values of predator's leaving rate (ρ), initial assets (I), and effects of intruder's time on reproductive success (c). The x-axis is in logarithmic scale to facilitate comparisons.

3.4.2.1 Optimal time to reemerge

Optimal time to reemerge (t_r^*) as a function of time of predator attack (t_p) is shown in Fig. 3.3 (p. 66) and 3.4 (p. 67) for an intruder with log-normal and exponential return times, respectively. One major difference between the exponential and the log-normal cases is that in the log-normal case there is an initial decrease in t_r^* as the time between the end of the aggressive interaction and the predator's attack increases. In other words, with a log-normal distribution of return times we can obtain a counterintuitive intensification of the effects of a past aggressive interaction

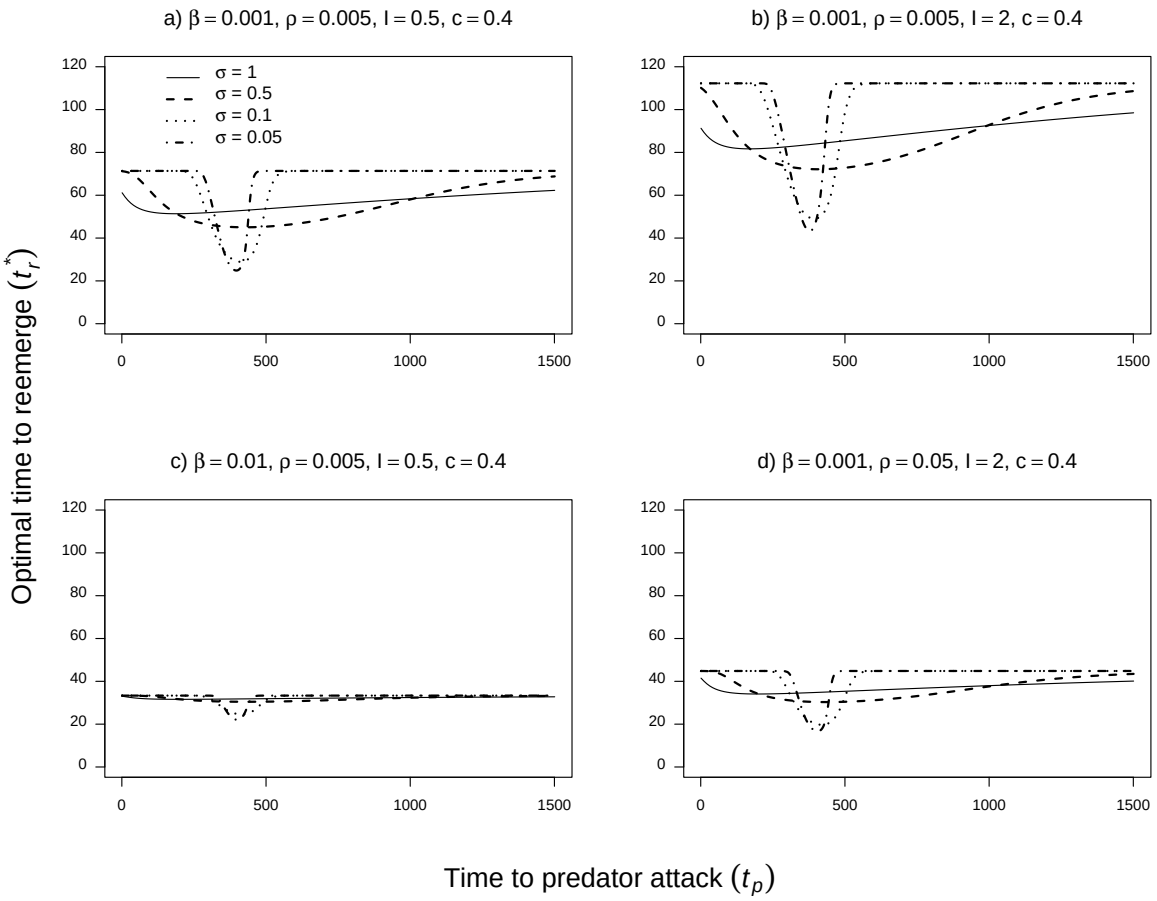

Figure 3.3: Optimal time to reemerge (t_r^*) as a function of time to predator attack (t_p), when time to intruder's return (conditional on reintruder attempting return) is log-normally distributed (with mean of log(return time) = 400). For explanation of other parameters see Table 3.1 (p. 62).

with time: t_r^* decreases with increasing t_p for values of t_p smaller than the maximum of the pdf (about 400). In addition the range of t_p's that exhibit an intuitive wearing-off of the effects of a past aggressive interaction (increase in t_r^* with increasing t_p) can be small compared to the t_p's that exhibit counterintuitive behaviour. In contrast, if intruders' return time follows an exponential distribution (or, more generally, a pdf with maximum value at 0 and monotonically decreasing thereafter), we cannot observe a counterintuitive intensification of the effects of a past aggressive interaction with increasing time to predator attack: the plot for the exponential

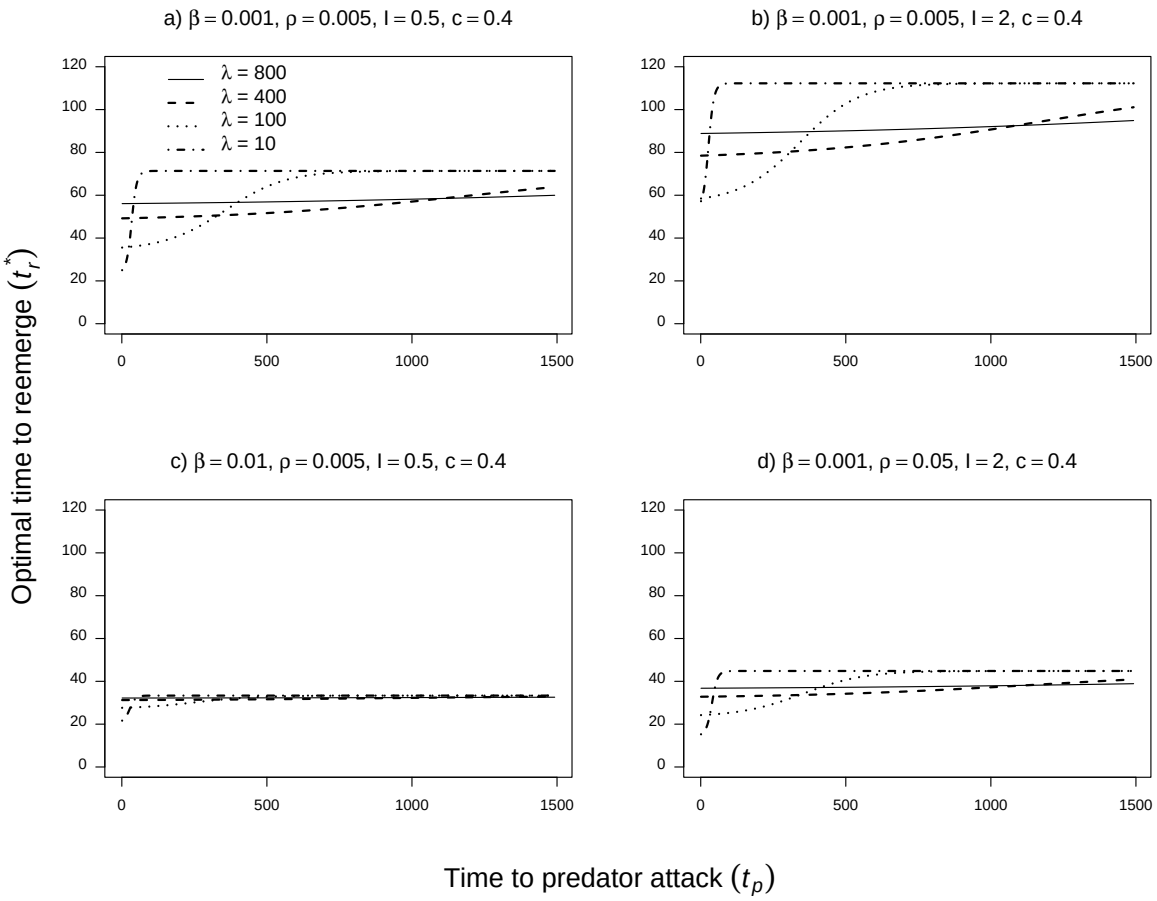

Figure 3.4: Optimal time to reemerge (t_r^*) as a function of time to predator attack (t_p), when time to intruder's return (conditional on reintruder attempting return) is exponentially distributed (with mean λ). For explanation of other parameters see Table 3.1 (p. 62).

case is like the plot for the log-normal case starting at $t_p \simeq 400$ (i.e., to the right of the maximum value of the pdf of the log-normal). The explanation of this pattern is the following: when the maximum of the pdf is some $t > 0$, as the time between t_p and that t increases (either because $t_p \ll t$ or $t_p \gg t$) the risk of a reintrusion in the near future decreases. In other words, if the predator attacks a long time before the maximum of that pdf, the resident need not worry about a particularly high risk of reintrusion for some time. Therefore, it is necessary to understand, at least qualitatively, the pattern of reintruder's return to make predictions about changes in reemergence time with variation in time to predator attack.

Increasing β (e.g., compare a) and c) in Fig. 3.3 and 3.4) decreases the effects of the reintruder: the relative importance of the reintrusion becomes smaller as the number of other intruders increases, because other conspecifics (and not the reintruder) are the major threat. Decreasing the probability of reintrusion also decreases the effects of the reintruder: the smallest possible t_r^* is larger, and t_r^* reaches the plateau faster (i.e., at smaller t_p). Likewise, increasing the speed at which reproductive success decreases with intruders' time in the territory (i.e., increasing c) or decreasing initial assets (i.e.,decreasing I —e.g., compare a) and b) in Fig. 3.3 and 3.4) decreases the effect of the reintruder: for any given β, faster loss of fitness with intrusion (or smaller initial reserves) decreases the maximum attainable difference in t_r^* (as the t_r^* in the absence of the reintruder is already small because the high rate of intrusion of other conspecifics forces the resident to reemerge sooner); however, the relative change (or, equivalently, the difference of the logarithms of time to reemerge) is sometimes larger and sometimes smaller with smaller I.

The variance of the return time of the reintruder has a strong effect on t_r^*. If the variance is small and the predator comes around the time when the pdf of the reintruder return is largest (e.g., 400; see Fig. 3.1a, $\sigma = 0.05$, p. 63) the probability of the reintruder returning in the near future is very high, and thus the effects on t_r^* are strong. If the predator comes long before that time, it is unlikely that the reintruder will return before the resident has already reemerged to prevent intrusions from other conspecifics (e.g., in fig. 3.3a, p. 66, with $\sigma = 400$, at $t_p \simeq 200, t_r^* = 70$; thus, the resident is reemerging at around 270, but the reintruder is unlikely to come long before 400). If the predator comes some time after the maximum of the pdf, it is unlikely that the reintruder will ever come, given that it has not come by that time. In contrast, with high variance, the probability of the reintruder coming in any particular interval is smaller, but this probability is spread over a larger time period (e.g., Fig. 3.1, p. 63) and even for large t_p's the probability is still high that the reintruder will come, given that it has not come by that time. Thus, with larger variances (i.e., less predictable reintruder), the effects of the reintruder in reemergence time can be observed for a larger range of t_p. In summary, with

small variances effects of a past aggressive interaction are more intense, but might be observable only for a small range of t_p's.

3.4.2.2 Optimal time to hide

When the reintruder has a stochastic reintrusion time, delaying hiding is never optimal, except for extremely small variances (e.g., Fig. 3.5) and low rate of intrusion of other intruders and only over a very small range of times to predator attack. To make delaying hiding optimal, the decrease in territorial costs and information gain has to be large enough to compensate the fast increase in the risk of mortality from delaying hiding. This can only be achieved if (1) there is almost certainty about the reintruder's return (variance close to zero — the estimate of the probability of reintrusion is updated using Bayes theorem [eq. 3.A.2, p. 80] and thus with small variances delaying hiding can provide a lot of information about the future probability of the reintruder's return) and (2) the loss of reproductive success from a reintruder has a major effect on fitness (e.g., when the rate of intrusion of other conspecifics is very low and initial territorial assets are small).

3.5 Discussion

This paper shows that risk of intrusion of conspecifics can have large effects on some components of the antipredator strategy: increased intrusion risk results in a decrease in time until reemergence from a refuge. When there is no threat from a reintruder but only risk of intrusion from other conspecifics, the optimal strategy (e.g., Fig. 3.2, p. 65) is to hide as soon as the predator attacks (i.e., not to delay hiding) and to modify time to reemerge as a function of the threat of invasion (larger numbers of intruders result in shorter reemergence) and initial resources (the higher the value of initial resources, the later an animal can afford to reemerge, as predicted from the asset-protection principle —Clark, 1994). The main focus of this paper

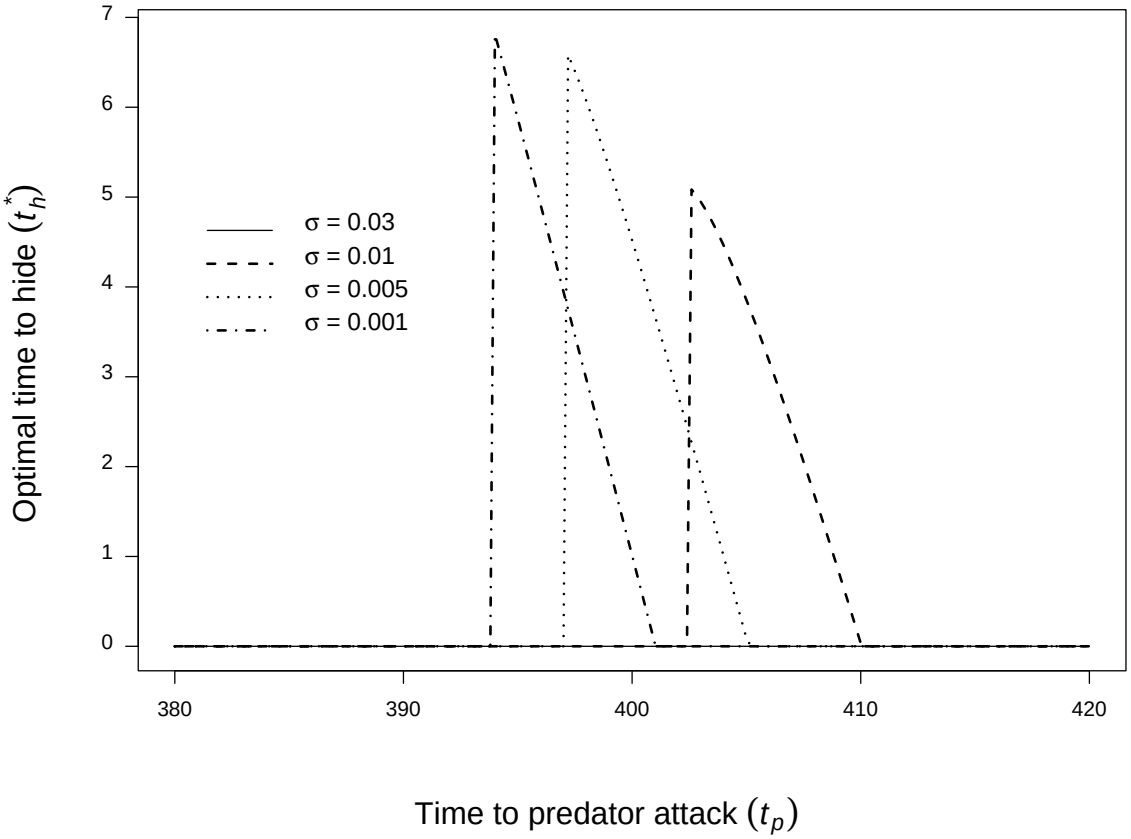

Figure 3.5: Examples of optimal time to hide (t_h^*), as a function of time to predator attack (t_p), when time to intruder's return (conditional on reintruder attempting return) is log-normally distributed (with mean of log(return time) = 400, $\beta = 0.001$, $\rho = 0.05$, $c = 0.4$, $I = 0.5$).

are the effects of a past aggressive interaction when intruders are persistent. In the presence of a reintruder, as was the case in the absence of a reintruder, the optimal strategy almost always involves hiding as soon as the predator attacks . However, reemergence time can be strongly affected by the possibility of a conspecific reintrusion. The extent of these effects will be modified by the time of the predator's attack and the behaviour of the reintruder (e.g., Fig. 3.3, p. 66 and 3.4, p. 67). Timing of attack of the predator and behaviour of the reintruder play a key role because the increase in territorial costs of intrusion is a consequence of a transient increase in the probability of reintrusion. As this probability increases, behaviour is modified

(earlier reemergence) at the expense of increased mortality risk, but it eventually returns to the same levels as in the absence of reintrusion.

3.5.1 Why not to delay hiding

Flight initiation behaviour (measured either as time to hide or approach or flight distance) has been shown empirically to respond to variation in predation risk (e.g., Bauwens & Thoen, 1981; Bulova, 1994; Cooper,1997; see review in Lima & Dill, 1990; Lima, 1998), but few studies have examined the effects of non-predatory factors such as increased cost of flight (Lima, 1998). Most evidence of delayed hiding with higher costs of hiding is limited to a few cases related to foraging costs of flight (see Lima, 1998, p. 237; Ydenberg & Dill, 1986, pp. 237-239). Recent empirical work has documented delayed hiding in mate guarding males (Cooper, 1997, 1999) and animals involved in ongoing aggressive interactions (Brick, 1998; Cooper, 1999; Díaz-Uriarte, 1999, experiment 2; Jakobsson et al., 1995). In addition, the model of Ydenberg & Dill (1986) predicts that time to initiate flight should increase with increasing cost of flight.

However, delaying hiding is rarely optimal in this model, which agrees with the empirical results of Díaz-Uriarte (1999) where male *Tropidurus hispidus* do not increase time to initiate escape if a predator attacks 5 min after an intruder is evicted from their territory; the predictions of this model, however, do not agree with the results of Díaz-Uriarte & Marler (in prep.) where there is also an increase in the delay to hide. In this model, delaying hiding can affect intrusion in two ways. First, delaying hiding prevents the invasion of both other conspecifics and the reintruder because while the resident is out of the refuge the intruders cannot successfully invade the territory. Second, delaying hiding serves to gain information about the reintruder's probability of return based upon the reintruder not having attempted to reinvade by the time the resident goes into hiding (the probability of reintrusion is updated using Bayes theorem —see expression 3.A.2, p. 80; no information can be gained about the other conspecifics, as the probability of invasion by other conspecifics is independent of past invasions). Information about

the reintruder's probability of return is valuable if it can modify future behaviour (Stephens, 1989; also Mangel, 1990), such as reemergence time. If the resident hides with a new estimate of the probability of future reintrusion very close to zero, time to reemerge could be much longer, therefore decreasing mortality at reemergence. Nevertheless, in most cases neither the decrease in intrusion costs nor the gain of information about the intruder's probability of return justify delaying hiding. These result depend on the attack of the predator being generally a very fast event, so that the small decrease in intrusion costs and/or the added information about the reintruder's likely behaviour cannot compensate the fast increase in mortality risk that results from delaying hiding.

The particular parameter values and functions used in this model affect the numerical results but do not change the qualitative conclusions. The main qualitative results only depend on, (1) that survivorship be a monotonically decreasing function of time to hide and (2) that the rate of decrease in survivorship with time to hide be faster than the rate of information gain (itself a function of the variance of reintruders' return). Both conditions are likely to hold in most biological systems.

What, then, explains the differences between the predictions of my model and those from the model of Ydenberg & Dill (1986) and the empirical findings of Brick (1998), Cooper (1999), Díaz-Uriarte (1999; experiment 2), Díaz-Uriarte & Marler (in prep.) and Jakobsson et al. (1995)? On the one hand, in Ydenberg & Dill's (1986) model there is always a cost to fleeing from predators (for example, loosing a very profitable prey item); in my model, the cost does not arise from fleeing itself but from hiding (which also explains why, in my model, when there is no reintruder delaying hiding can never be optimal). On the other hand, all of the empirical evidence, except for Díaz-Uriarte & Marler (in prep.), deals with animals actively engaged in a fight. In those situations the animals are facing an actual intrusion, and not just risk of a probable intrusion sometime in the future; when the animal is engaged in an ongoing fight fleeing itself (and not just hiding time) has a cost, as in the model of Ydenberg & Dill (1986), and this

cost could be much higher if the approaching predator is not an attacking one (Díaz-Uriarte, 1999).

Nevertheless, there are other costs of hiding soon such as interrupting foraging (e.g., Ydenberg & Dill, 1986; Lima, 1998) and degrading information acquisition (interrupting sampling —e.g., Dall et al., 1999) that have not been considered in this model. Moreover, these costs could be comparatively high if the approaching predator is not an attacking one (Díaz-Uriarte, 1999; see also Lima & Dill, 1990), whereas in the present model the approaching predator was always attacking. Finally, delaying hiding when there is uncertainty about the predator's intentions (attacking vs. non-attacking) could actually provide information about the probability that the approaching predator is an attacking one and thus modify, for example, reemergence time. These effects are currently under investigation. But the main conclusion from my model regarding flight behaviour is that the risk of a potential intrusion, per se, will very rarely justify delaying hiding from an attacking predator. Interestingly, in the experiments in Díaz-Uriarte & Marler (in prep.) the predator's approach speed was about half of the predator's approach speed in Díaz-Uriarte (1999), and thus makes more likely that these additional costs of hiding could be detected. In summary, the differences with the model of Ydenberg & Dill (1986) suggest that trade-offs between predation and foraging could be very different from those between predation and territorial defence. Whereas in the former it is interrupting foraging that is most costly, in the latter costs arising from hiding and interruption of information acquisition could be the most relevant.

3.5.2 Using multiple responses to characterise antipredator behaviour, and applying and testing the model

The above results have been obtained because we have characterised antipredator behaviour using two variables, time to hide and time to reemerge, instead of a single one (such as proportion of time hiding). As emphasised by Lima & Dill (1990), in the study of conflicting demands

of antipredator behaviour it is necessary to identify the key behavioural decisions involved in predator avoidance; this context specificity is a necessary step to guide further empirical work and generate testable predictions.

The results of this paper also show that applying and extending this model requires a better understanding of reintrusion patterns in nature, since the re-invasion behaviour of the reintruder can have a large effect on the detectability of effects of a past aggressive interaction and the type of change of time to reemerge with variation in time to predator attack. Unfortunately, there is no information about reintrusion patterns in nature. A pdf of return times with a maximum not at zero creates two potential problems for empirical work. First, there will be a window of times to predator attack during which increasing the time between the end of the eviction of the intruder and the predator attack results in a counterintuitive increase in the effects of the past aggressive interaction (as the time to reemerge decreases —Fig. 3.3, p. 66). Second, and more importantly, the largest effects will be detected around the (generally unknown) maximum of the pdf, but might be negligible shortly after the intruder is evicted (e.g., Fig. 3.3, p. 66). This is not a problem if the reintruders return as with an exponential distribution (or, more generally, a pdf with maximum value at 0 and monotonically decreasing thereafter); in this case, the best way to detect an effect of past aggressive interactions is to expose the resident to a simulated predator attack shortly after the resident has evicted a reintruder (Fig. 3.4, p. 67).

An increase in predation exposure following an aggressive encounter emphasises that a similar functional explanation, adaptive response by a territorial resident to a transient increase in the probability of intrusion, could underlie different behavioural phenomena: past aggressive interactions are known to increase the time invested in territorial vigilance (e.g., great tits: Ydenberg & Krebs, 1987; Kacelnick et al., 1981) and the frequency of territorial displays (e.g., the lizards *Sceloporus jarrovi* and *Urosaurus ornatus*; Moore, 1987; Thompson & Moore, 1992), and in a wide range of taxa (e.g., Adamo & Hoy, 1995; Chase et al., 1994) past experiences of victory make winning future encounters more likely. In addition, the consequences of past ag-

gressive interactions are receiving increased theoretical attention (e.g., Johnstone & Dugatkin, 2000), but this is, to my knowledge, the first theoretical work to relate past aggressive interactions with antipredator behaviour. Given the potentially far-reaching consequences of these effects, and their connections to other behavioural and ecological phenomena, it is expected that the present paper will spur further theoretical and empirical work.

3.6 Acknowledgements

My thinking about this problem has been shaped by several years of discussion with C. Lázaro-Perea. I have also benefited from suggestions from E. Davis, K. Gross, and specially A. R. Ives and E. V. Nordheim. K. Gross, P. Hoff, A. R. Ives, E. V. Nordheim, and specially T. Kurtz answered questions about the math. A. R. Ives, C. Lázaro-Perea and C. A. Marler provided many useful comments on the ms. Work on the model and manuscript were greatly facilitated by the programs R and Lyx (both open-source and relased under the GNU GPL license). Manuscript preparation partially supported by a John and Virginia Emlen Graduate Student Fellowship, Department of Zoology, University of Wisconsin-Madison.

3.7 References

Adamo, S. A. & Hoy, R. R. 1995. Agonistic behavior in male and female field crickets, Grillus bimaculatus, and how behavioral context influences its expression. Anim. Behav., 49, 1491-1501.

Baker, R. L., Elkin, C. M., & Brennan, H. A. 1999. Aggressive interactions and risk of fish predation for larval damselflies. J. Insect Behav., 12, 213-223.

Bauwens, D. & Thoen, C. 1981. Escape tactics and vulnerability to predation associated with reproduction in the lizard Lacerta vivipara. J. Anim. Ecol., 50, 733-743.

Brick, O. 1998. Fighting behaviour, vigilance and predation risk in the cichlid fish Nannacara anomala. Anim. Behav., 56, 309-317.

Brick, O. 1999. A test of the sequential assessment game: the effect of increased cost of sampling. Behav. Ecol., 10, 726-732.

Bulova, S. J. 1994. Ecological correlates of population and individual variation in antipredator behavior of two species of desert lizards. Copeia, 1994, 980-992.

Chase, I. D., Bartolomeo, C., & Dugatkin, L. A. 1994. Aggressive interactions and inter-contest interval: how long do winners keep winning? Anim. Behav., 48, 393-400.

Clark, C. W. 1994. Antipredator behavior and the asset-protection principle. Behav. Ecol., 5, 159-170.

Cooper, W. E. Jr. 1997. Factors affecting risk and cost of escape by the broad-headed skink (Eumeces laticeps): predator speed, directness of approach, and female presence. Herpetologica, 53, 464-474.

Cooper, W. E. Jr. 1999. Tradeoffs between courtship, fighting, and antipredatory behavior by a lizard, Eumeces laticeps. Behav. Ecol. Sociobiol., 47, 54-59.

Dall, S. R. X., McNamara, J. M., & Cuthill, I. C. 1999. Interruptions to foraging and learning in a changing environment. Anim. Behav., 57, 233-241.

Díaz-Uriarte, R. 1999. Anti-predator behaviour changes following an aggressive encounter in the lizard Tropidurus hispidus. Proc. R. Soc. Lond. B., 266, 2457-2464.

Jakobsson, S., Brick, O., & Kullberg, C. 1995. Escalated fighting behaviour incurs increased predation risk. Anim. Behav., 49, 235-239.

Johnstone, R. A. & Dugatkin, L. A. 2000. Coalition formation in animals and the nature of winner and looser effects. Proc. R. Soc. Lond. B., 267, 17-21.

Kacelnick, A., Houston, A. I., & Krebs, J. R. 1981. Optimal foraging and territorial defence in the great tit (Parus major). Behav. Ecol. Sociobiol., 8, 35-40.

Krupa, J. J. and Sih, A. 1998. Fishing spiders, green sunfish, and a stream-dwelling waterstrider: male-female conflict and prey responses to single versus multiple predator environments. Oecologia, 117, 258-265.

Lima, S. L. 1998. Stress and decision making under the risk of predation: recent developments from behavioral, reproductive, and ecological perspectives. In: Advances in the study of behavior, v. 27. (Ed. by P. J. B. Slater, J. S. Rosenblat, C. T. Snowdon, & M. Milinski), pp. 215-290. San Diego: Academic Press

Lima, S. L. & Dill, L. 1990. Behavioral decisions made under the risk of predation: a review and prospectus. Can. J. Zool., 68, 619-640.

Mangel, M. 1990. Dynamic information in uncertain and changing worlds. J. Theor. Biol., 146, 317-332.

Martel, G. 1996. Growth rate and influence of predation risk on territoriality in juvenile coho salmon (Oncorhynchus kisutch). Can. J. Fish. Aquat. Sci., 53, 660-669.

Martín, J. & López, p. 1999. When to come out from a refuge: risk-sensitive and state-dependent decisions in an alpine lizard. Behav. Ecol., 10, 487-492.

Moore, M. C. 1987. Circulating steroid hormones during rapid aggressive responses of territorial male mountain spiny lizards, Sceloporus jarrovi. Horm. Behav., 21, 511-521.

Roussas, G. R. 1997. A course in mathematical statistics. San Diego: Academic Press.

Schwarzkopf, L. & Shine, R. 1992. Costs of reproduction in lizards - escape tactics and susceptibility to predation. Behav. Ecol. Sociobiol., 31, 17-25.

Stamps, J. A. & Krishnan, V. V. 1995. Territory acquisition in lizards: III. Competing for space . Anim. Behav., 49, 679-693.

Stamps. J. A. & Krishnan, V. V. 1998. Territory acquisition in lizards. IV. Obtaining high status and exclusive home ranges. Anim. Behav., 55, 461-472.

Stephens, D. W. 1989. Variance and the value of information. Am. Nat., 134, 128-140.

Thompson, C. W. & Moore, M. C. 1992. Behavioral and hormonal correlates of alternative reproductive strategies in a polygynous lizard: tests of the relative plasticity and challenge hypotheses. Horm. Behav., 26, 568-585.

Whitehouse, M. E. A. 1997. The befits of stealing from a predator: foraging rates, predation risk, and intraspecific aggression in the kleptoparasitic spider Argyrodes antipodiana. Behav. Ecol., 8, 663-667.

Williams, D. 1991. Probability With Martingales. Cambridge: Cambridge University Press.

Wisenden, B. D. & Sargent, R. C. 1997. Antipredator behaviour and suppressed aggression by convict cichlids in response to injury-released chemical cues of conspecifics but not to those of an allopatric heterospecific. Ethology, 103, 283-291.

Ydenberg, R. C. & Dill, L. M. 1986. The economics of fleeing from predators. In: Advances in the study of behavior, v. 16. (Ed. by J. S. Rosenblatt, C. Beer, M.-C. Busnel, & P. J. B. Slater), pp. 229-249. New York: Academic Press

Ydenberg, R. C. & Krebs, J. R. 1987. The tradeoff between territorial defense and foraging in the great tit (Parus major). Amer. Zool., 27, 337-346.

Appendix 3.A: expected accumulated time that intruders spend in the territory

3.A.1 Time spent by other conspecifics

If the presence of a predator does not affect the entry of intruders, the expected total time that the other conspecifics accumulate within a territory in the absence of the territory owner is given by

$$E\left[\int_h^r N(t)\,\mathrm{d}t\right] = \int_h^r E\left[N(t)\right]\,\mathrm{d}t = \int_h^r \beta(t-h)\,\mathrm{d}t = \beta\frac{(r-h)^2}{2} \qquad \text{(3.A-1)}$$

where $N(t)$ is the number of other conspecifics by time t, h is the time at which the animal hides (i.e., $t_p + t_h$), and r is the time at which the animal reemerges (i.e., $t_p + t_h + t_r$); the first equal sign (interchange of order of integration and expectation) follows from Fubini's theorem (e.g., Williams, 1991, ch. 8) and the second results from direct substitution of the expected value of a Poisson random variable. Once intruders are present in the territory they no longer leave.

Expression 3.A-1 needs to be modified because no intruder can enter the territory while the predator is in the area; therefore, the starting time of the process is not h, but a random variable, z, whose pdf is the pdf of the time at which the predator leaves the area (i.e., $f_Z(z) = \rho \mathrm{e}^{-\rho(z-h)}$). Then, using conditional expectation ($E[Y] = E[E[Y|Z]] = \int E[Y|Z=z] f_Z(z)\,\mathrm{d}z$) the expected total time that the other conspecifics accumulate is given by

$$\int_h^r \beta\frac{(r-u)^2}{2}\rho\,\mathrm{e}^{-\rho(u-h)}\,\mathrm{d}u = \int_0^{t_r} \beta\frac{(t_r-s)^2}{2}\rho\,\mathrm{e}^{-\rho s}\,\mathrm{d}s = \frac{\beta}{\rho^2} - \frac{\beta}{\rho^2\mathrm{e}^{\rho t_r}} - \frac{\beta t_r}{\rho} + \frac{\beta t_r^2}{2}. \qquad \text{(3.A-2)}$$

3.A.2 Time spent by the reintruder

First, suppose that, conditional on the reintruder attempting a return, the reintruder return time is fixed (i.e., the pdf of t is 1 for $t = t_i$ and 0 otherwise). Define $t_{ip} = t_i - t_p$ as the time at which the intruder returns with respect to the predator attack. Since the reintruder can only invade successfully if the resident is hiding and the predator is not present, the expected time spent by the reintruder is given by

$$(r - t_i)(1 - \mathrm{e}^{-\rho(t_i-(t_p+t_h))})p = (t_r + t_h + t_{pi})(1 - \mathrm{e}^{-\rho(t_{ip}-t_h)})p, \tag{3.A-3}$$

whenever $t_h < t_{ip} < (t_r + t_h)$, and 0 otherwise.

If time when the reintruder attempts to return, t_i, has a pdf $f_T(t_i)$, then the expected time that the reintruder spends in the territory can be found as follows. The random variable of interest is not t_i but the time that the reintruder spends in the territory, given by $r - t_i$. Define a random variable X that takes the value $r - t_i$ when the reintruder successfully reinvades, and 0 otherwise (i.e., if the reintruder never attempts to return, or if it attempts to return while the resident is hiding —between r and h— but is unsuccessful because the predator is present), so $0 \leq x \leq r - t_i$. We are interested in the expected value of X conditional on the reintruder not having attempted a return by $h = t_p + t_h$. The expectation can be written as

$$\begin{aligned} &E[X|\text{No attempted invasion by } h] = \\ &E[X|(\text{No attempted invasion by } h) \cap (\text{Attempted invasion})] \\ &P\text{Attempted invasion}|\text{No attempted invasion by } h\ . \end{aligned} \tag{3.A-4}$$

Eq. 3.A-4 comes from the relationship

$$E[X|A] = E[X|A \cap B]\, P[B|A] + E[X|A \cap B^c]\, P[B^c|A], \tag{3.A-5}$$

where X is a random variable and A and B are events or sets, $\cap$ denotes intersection of events, and ^{c} denotes the complement. To derive eq. 3.A-4 from eq. 3.A-5 note that X takes value 0 when no attempted invasion, or

$$E[X|(\text{No attempted invasion by } h) \cap (\text{No attempted invasion})] = 0.$$

To evaluate eq. 3.A-4 we will need

$$P\{\text{No invasion by } h\} = (1-p) + p(1 - F_T(h)) = 1 - pF_T(h)$$

where $F_T(t)$ is the cumulative distribution function of time to reintrusion. Thus,

$$\begin{aligned} &P\{\text{Attempted invasion}|\text{No attempted invasion by } h\} = \\ &1 - P\{\text{No attempted invasion}|\text{No attempted invasion by } h\} = \\ \text{(from Bayes theorem)} \qquad &1 - \frac{1-p}{1-pF_T(h)} = \frac{p(1-F_T(h))}{1-pF_T(h)}. \end{aligned} \tag{3.A-6}$$

We need to obtain the pdf $f_{(X|(\text{No attempted invasion by } h)\cap(\text{Attempted invasion}))}(x)$ to compute the expectation in (3.A-4). In what follows I only show the pdf for $0 < x \leq r-h$, because when $x = 0$ it does not contribute to the expectation; in this interval $f_X(x) = f_T(r-x)$ (e.g., Roussas, 1997, pp. 215 & ff.). Hence, for $0 < x \leq r - h$ or, equivalently, $0 < x \leq t_r$, and using the definition of conditional pdf (e.g., Roussas, 1997, pp. 93 & ff.),

$$f_{(X|(\text{No attempted invasion by } h)\cap(\text{Attempted invasion}))}(x) = \frac{pf_T(r-x)}{p(1-F_T(h))} P\{\text{No predator at } r-x\}; \tag{3.A-7}$$

where

$$P\{\text{No predator at } r-x\} = 1 - \mathrm{e}^{-\rho(r-x-h)} = 1 - \mathrm{e}^{-\rho(t_r-x)}, \tag{3.A-8}$$

from expression 3.2 and since the process of the predator leaving starts at the time the resident hides (h). Finally, substituting (3.A-8) into (3.A-7), using (3.A.2) in (3.A-4), applying the definition of expectation to the random variable in (3.A-7), and simplifying and showing results in terms of t_h and t_r, we obtain

$$E[X|\text{No attempted invasion by } t_{p+}t_h] = \frac{p}{1-pF_T(t_p+t_h)} \int_0^{t_r} x f_T(t_p+t_h+t_r-x)\left(1-\mathrm{e}^{-\rho(t_r-x)}\right) \mathrm{d}x. \tag{3.A-9}$$

In all the figures shown in this paper, I evaluated this integral using numerical integration.

Appendix 3.B: Results for reintruders with fixed reintrusion time

This appendix shows the results for optimal time to hide and optimal time to reemerge when the reintruder has a fixed time of return. These results are similar to those that we can obtain for a stochastic intruder with variance of return time almost zero. To make these results comparable to those of stochastic reintruders, I have set the time of return at 400. The main difference between these results and those from a stochastic intruder are that, in this case, we can appreciate the effects of the predator precluding the reintruder's return.

3.B.1 Optimal time to reemerge

Fig. 3.6 (p. 83) shows optimal time to reemerge, t_r^*, as a function of t_p for different combinations of β, ρ , I, and c when $t_i = 400$. To explain the results I will refer to two points in Fig. 3.6, t_1 and t_2 that divide the range of t_p into three distinct regions, and are the t_p's that correspond to the minimum and maximum t_r^*. A $t_p > 400$ means that the predator is initiating its attack after the intruder is scheduled to come and thus t_r^* is the same as if there were no reintruder. If the reintruder comes shortly after the predator attacks ($t_1 < t_p < 400$) t_r^* is large: it is very unlikely that the reintruder will invade the territory (as that can only happen if the predator is no longer present), and thus the resident can reemerge late; for example, with decreasing ρ the predator is likely to stay longer, which results in larger t_r^* at t_p close to 400 —see Fig. 3.6b vs. 3.6a. For $t_2 < t_p < t_1$, t_r^* decreases linearly with t_p: the resident is reemerging at t_i ($t_r^* = t_{ip} = 400 - t_p$) so that the reintruder does not accumulate any time in the territory. For $t_p < t_2$, t_r^* is not affected by changes in t_p: to prevent further increases in territorial costs from the other conspecifics' intrusions the resident is reemerging before the reintruder is scheduled to come, and t_r^* is the same as if there were no reintruder.

Increasing β increases the number of conspecifics that can intrude per unit time, and decreases the sensitivity of t_r^* to changes in t_p, because the effect of the reintruder decreases

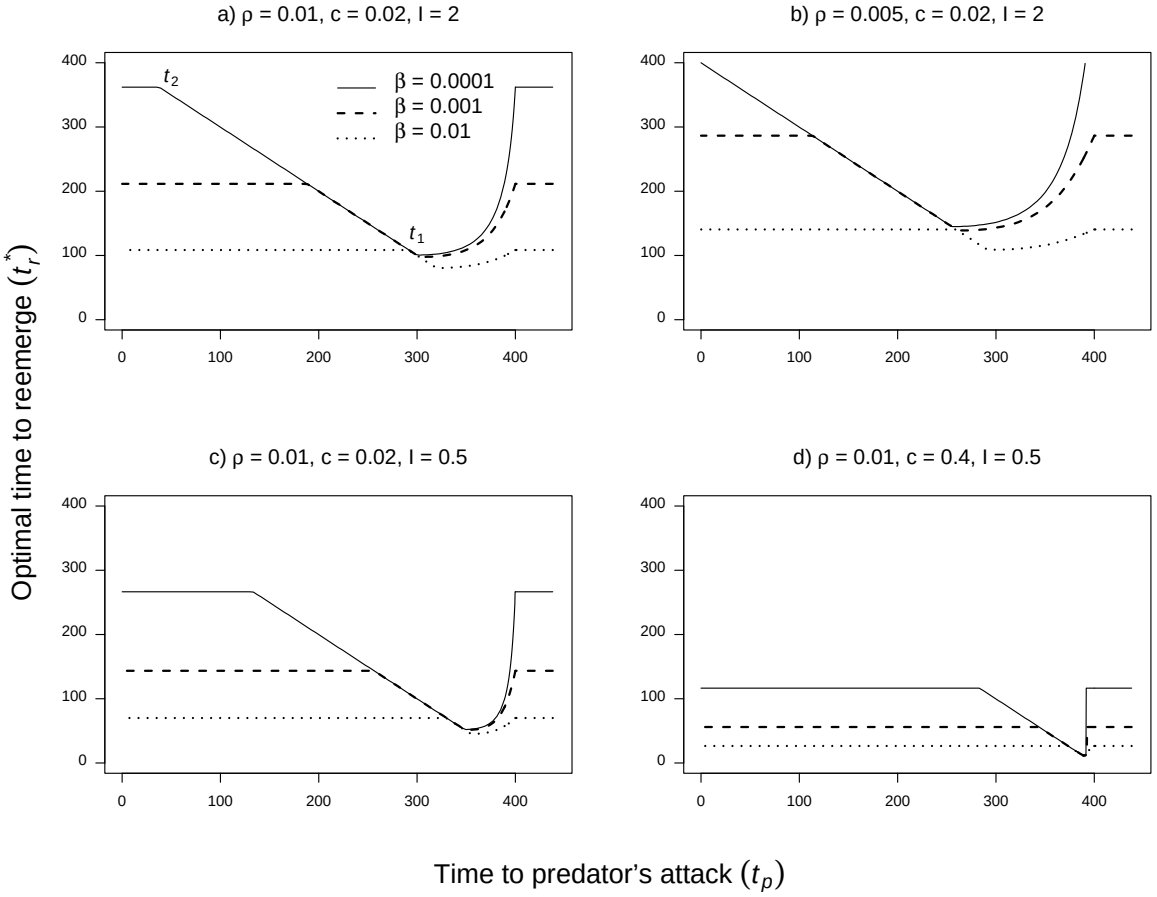

Figure 3.6: Optimal time to reemerge (t_r^*) as a function of time to predator attack (t_p), when time to reintruder's return (t_i) is 400. Points t_1 and t_2 (panel c) divide the range of t_p into three regions: when $t_1 < t_p < 400$ t_r^* increases as t_p increases; for $t_2 < t_p < t_1$ $t_r^* = 400 - t_p$; for $t_p < t_2$ the behaviour of the resident is insensitive to the past aggressive interaction (t_r^* does not depend on t_p). Values of $t_p > 400$ correspond to the predator attacking after the reintruder is scheduled to come, and thus t_r^* is the same as in the absence of a reintruder (i.e., there are no effects of reintrusion).

relative to other conspecifics. The largest possible difference in t_r^* (between points t_1 and t_2) is smaller because t_2 is shifted to the right; in other words, as we increase β the t_p at which the resident's behaviour is no longer affected by the reintruder is larger. Decreasing I also decreases sensitivity to the reintruder (Fig. 3.6a vs. 3.6c) as does increasing c (Fig. 3.6c vs. 3.6d): if initial assets are small or loss of reproductive success fast, the reproductive success that a resident can afford to loose to intrusion decreases; this causes the maximum t_r^* to decrease: t_2 is shifted

to the right and this is not compensated by the small decrease in t_r^* at t_1. However, changes in I and c do not make the reintruder less important relative to the other conspecifics: they simply magnify the effect of any territorial losses. Finally, increasing t_p (i.e., staging a predator attack a longer time after an intruder is chased away) will decrease t_r^* whenever $t_2 < t_p < t_1$; this is counterintuitive, because the effect of a past aggressive interaction becomes stronger (t_r^* smaller compared to a situation without reintruder) as the predator attacks a longer time after the intruder was chased away. The cause of this counterintuitive result is different from the counterintuitive result for a reintruder with log-normal return time shown in Fig. 3.3 (p. 66). Finally, the "intuitive" result of a wearing-off of the effects of a past aggressive interaction as t_p increases is only observed for $400 < t_p < t_1$, but this region ($400 < t_p < t_1$) might be small.

3.B.2 Optimal time to hide

With a reintruder that returns at a fixed time (t_i), optimal time to hide, t_h^*, can only take two values: 0 and t_{ip} (the time at which reintruder attempts to return relative to predator's attack). When t_h is 0, the resident avoids mortality risks from predation during the initial attack. When $t_h = t_{ip}$ (i.e., delayed hiding) the resident prevents the reintruder from coming back (as the reintruder can only come back if $t_h < t_{ip} < (t_r + t_h)$). No other value of t_h can be optimal; any value of t_h between 0 and t_{ip} exposes the resident to predation without preventing the reintruder from returning, and values of $t_h > t_{ip}$ result in increases in mortality risk with respect to $t_h = t_{ip}$ with no further reduction in territorial intrusion risk. Delaying hiding will also allow the resident to reemerge later than if it had hid at 0 as the re-intruder is no longer a threat and reemergence is only dictated by the rate of intrusion of other intruders.

Fig. 3.7a (p. 85) shows t_h^* as a function of time to predator attack (t_p) when $t_i = 400$ for three different β's. In every case, when $t_p < 390$ then $t_{ip} > 10$ and thus t_h^* is always 0: delaying hiding in these cases would require delaying hiding for more than 10 time units, which results in no survivorship. When $t_p > 400$ the predator is attacking after the reintruder is scheduled

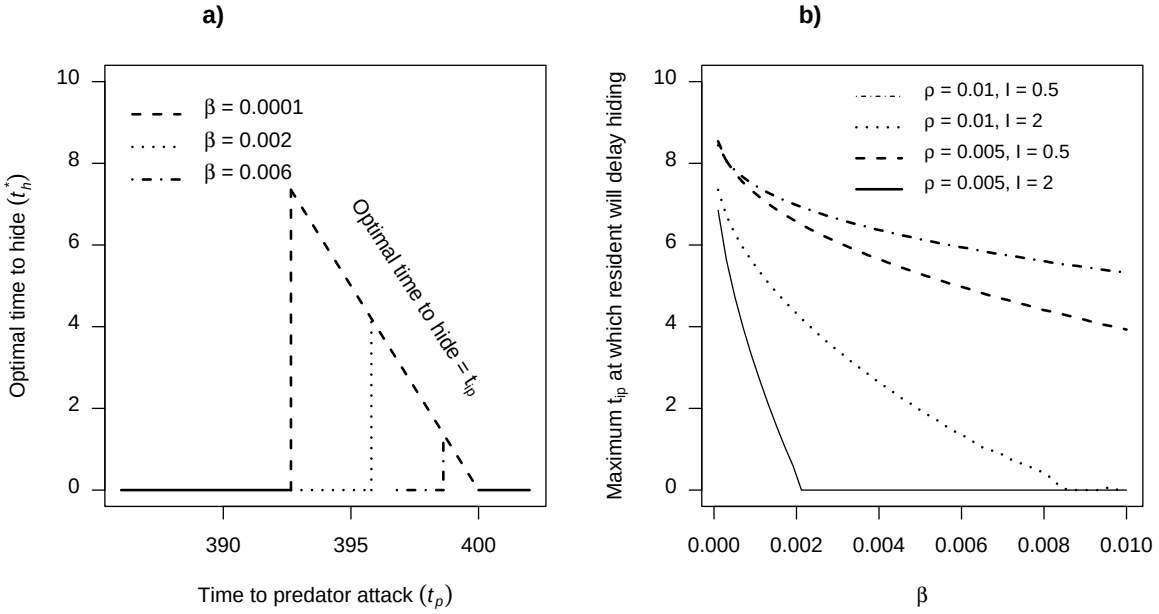

Figure 3.7: Optimal time to hide, t_h^*. a) Effects of time to predator attack (t_p) on t_h^* when the reintruder is scheduled to come at $t_i = 400$. In the case represented, for instance, when $\beta = 0.002$, the resident will delay hiding if $396 < t_p < 400$ (see text for explanation), and the delay will be equal to $t_{ip} = 400 - t_i$; at any other values the resident will hide immediately ($t_h = 0$). b) Maximum t_{ip} (time or reintruder's return relative to the attack of the predator) at which a resident will delay hiding, as a function of rate of intrusion of other conspecifics (β) for different values of predator's leaving rate (ρ) and initial assets (I). The value shown in the figure is the largest t_{ip} for which fitness is larger when $t_h = t_{ip}$ compared to $t_h = 0$. For any intruder returning at a t_{ip} below the line, the resident's optimal behaviour will be to make make $t_h = t_{ip}$; for any t_{ip} above the line the optimal t_h will be zero.

to come, so the reintruder is no longer a threat and thus t_h^* is 0. For $390 < t_p < 400$ it might be optimal to delay hiding; in this region t_h^* can be either 0 (no delayed hiding) or t_{ip}; thus, the line in Fig. 3.7a has a slope of -1 ($t_h^* = t_{ip} = 400 - t_p$). In general, it is more likely that delaying hiding will be optimal at small t_{ip}: here, delaying hiding does not represent a large increase in mortality, whereas for large t_{ip} the mortality risk of delaying hiding will be very

large). However, delaying hiding, if at all, will only be observed in a small range of values of t_p (when the predator attacks shortly before the reintruder is scheduled to come). Fig. 3.7a and b also shows the effects of β on $t_h^* = t_{ip}$. As the rate of intrusion of other conspecifics increases, the relevance of the reintruder decreases, and thus it becomes less worthy of increasing mortality risks. The optimal time to delay hiding depends also on the effect of intruders on reproductive success (c), the initial territorial assets (I), and the predator's behaviour (ρ) (Fig. 3.7b).

Chapter 4

Cross-over trials in animal behaviour. I: Misuse, carry-over effects, and design

4.1 Abstract

Cross-over trials (experiments where each experimental unit receives two or more treatments through time) are frequently used in animal behaviour studies as they allow experiments with relatively small numbers of subjects that nonetheless achieve high statistical power by using each subject as its own control. However, cross-over trials are often analyzed incorrectly in the behavioural literature; the major problems are failure to consider period and carry-over effects. In this chapter I first show these problems by using articles published in twelve issues of Animal Behaviour (July 1998 to June, 1999); 22 papers use crossover designs in at least one experiment, but because of potentially inappropriate analyses the conclusions in each of these papers are questionable. In addition, statistical textbooks frequently used by behaviourists either do not mention cross-over designs or provide potentially misleading advice. In this paper I explain why the usual analyses of cross-over trials (paired t-tests or non-parametric analogues) are often inappropriate, then discuss the problems associated with carry-over effects, and finally review the design of cross-over trials. If design and wash-out periods are given the appropriate consideration, cross-over designs can be very powerful tools for behaviourists whenever obtaining new subjects is more costly than repeatedly testing the same individual, and thus cross-over designs can be useful in particular for researchers working in the lab or in field enclosures where

animals require lengthy training or habituation

4.2 Introduction

In cross-over trials each experimental unit receives two or more treatments through time; in the simplest case of two treatments, the subject is first given one of the treatments and then crosses over to the other treatment (Jones and Kenward, 1989 –hereafter JK–; Ratkowsky et al., 1993 –hereafter REA–; Senn, 1993 a –hereafter SN–; Vonesh & Chinchilli, 1997). Thus, cross-over studies differ from parallel studies where each subject is exposed to the same treatment for the duration of the experiment. In cross-over trials at least one key covariate (treatment) changes within-subject over time. As the comparison of treatments is made within subjects, each subject acts as its own control which increases statistical power to detect a treatment effect (e.g., Crowder & Hand, 1990, p. 101; SN, pp. 201 & ff.). This is particularly important when repeated testing of one subject is much simpler than recruitment of new subjects. For these reasons, cross-over trials are frequently used in behavioural experiments.

However, cross-over trials are often analysed inappropriately, as if they were matched pairs or "typical" repeated-measures designs, which they are not. The main problems are, first, not accounting for period effects (which leads to the inappropriate use of paired t-tests in the two-treatment, two-period case) and, second, failure to consider carry-over effects. (A treatment effect is the effect of a treatment at the time of its application, whereas carry-over effects are effects of a treatment that persist after the end of the period, and a period is each one of the occasions in which a treatment is applied; see “Terminology”, p. 90.)

For instance, in the twelve issues of Animal Behaviour from July, 1998 to June, 1999, there are 22 articles that use cross-over designs in at least one experiment. Eight of these papers use variants of the two-treatment, two-period design (generally the typical 2x2 design); 17 papers use designs for more than two treatments. Results are analysed with paired t-tests or

Wilcoxon's signed-rank test for 2 treatment designs, or with linear models (usually referred to as "repeated measures ANOVA" or in some cases mixed-effects models), and on a few occasions with methods specific for categorical data. Only two studies explicitly consider period effects (order of presentation), and one mentions that there are "no effects of order of presentation" (although the test is not explained); but no paper explains how potential carry-over effects are dealt with. Counterbalancing (each treatment appears in each period the same number of times) is used in 11 papers. When counterbalancing is not used, order of presentation is "randomized." Thus, it seems that most authors believe that counterbalancing or "randomization" of order of presentation, per se, will take care of any other nuisances (periods and carry-overs); but, as we will see, this is not true. Authors seem unaware that carry-over effects can bias their conclusions. The practical consequences of the analyses used in these papers are that: a) if there are carry-over effects, all reported results could be biased; b) even in the absence of carry-over effects, in the studies that do not use counterbalancing the estimates of treatment effects are biased if there are period effects; c) in studies that use counterbalancing, the estimates of the variance of treatment effects are overestimated if there are period effects. Therefore, the conclusions reached in every one of these papers are questionable: the lack of effects reported in some studies could be the consequence of inflated variances, and the significant effects reported in other experiments could be the result of either period or carry-over effects.

Statistics textbooks used by behaviourists such as Colgan (1978), Lehner (1979), Bart et al. (1998), Bailey (1995), Campbell (1989), and Sokal & Rohlf (1995) do not mention cross-over designs. Other texts provide potentially misleading advice; Martin and Bateson (1993, p. 29-30) apparently would use a paired test to analyse a 2x2 design; Zar (1996, p. 259-263) analyses a cross-over design, and refers to carry-over, but he fails to mention that period should be incorporated in the analyses, and seems to imply that counterbalancing, per se, can eliminate problems from carry-over effects; Edgington (1995) suggests counterbalancing (pp. 114-117) to prevent undesired effects from order of presentation; Zolman (1993), although explicitly mentions cross-over designs and discusses carry-over effects (pp. 59-63), apparently suggests

that a paired t-test is appropriate for a 2x2 design (p. 160).

The 22 examples from one year of Animal Behaviour show that cross-overs, a powerful and widespread type of design, are often analysed inappropriately; and the textbook examples indicate that information on the appropriate design and analysis of cross-over trials is not accessible to animal behaviour researchers. Thus, the main objective of this paper is to make animal behaviour researchers (and reviewers) aware of the most important pitfalls in the design and analysis of cross-over trials. I first explain why the usual analyses of cross-over trials in animal behaviour research are inappropriate, then I discuss the problems of carry-over effects, next I review the design of cross-over trials, and I conclude with a discussion on when to use cross-over designs in behavioural ecology experiments. In a different paper (Díaz-Uriarte, in review —next chapter—; hereafter DU2) I review the statistical methods available for the analysis of data from cross-over experiments in animal behaviour research.

4.3 Terminology

Before we can understand the problems of some of the analyses of cross-over trials, we need to define a few terms. A direct treatment or simply treatment effect is the effect of a treatment at the time of its application. A period is each one of the occasions in which a treatment is applied. Carry-over effects are effects of a treatment that persist after the end of the treatment period; in other words, the response to a current treatment is affected by what treatment was applied in a previous period. A sequence is the order in which the within-individual treatments are applied. Designs will be referred to using sequences, such as ABB,BAA, which means that animals assigned to the sequence ABB are first given treatment A (1st period), then B (2nd period), then B (3rd period), and animals assigned to the BAA sequence are first given B, then A, then A (1st, 2nd, and 3rd periods, respectively). Designs are examined in detail later.

4.4 Example of the "usual" analyses and their problems

The 2x2 cross-over design (the design with sequences AB,BA) is frequently analysed using a paired t-test; this is equivalent to subtracting the response value under treatment B from the response value under treatment A for each individual and testing whether the mean is significantly different from 0 with a one sample t-test. However, in many behavioural experiments period has an effect: whether a response is measured on the first or second occasion, per se, will affect the value of the response (e.g., through habituation). With period effects the analysis above is inappropriate for two reasons (SN, p. 38; also Schneider, 1983). First, if there are unequal numbers of subjects in each sequence, the test and the estimate of treatment effects will be biased. (Bias means that the expected value of the estimator is not equal to the parameter we are trying to estimate; bias does not decrease with increasing sample size). Second, even if there are equal numbers of subjects in each sequence, we lose power: period is a systematic trend, but by lumping together animals from both sequences, we are ascribing this systematic variation to the random component (the error term) and the standard errors of our estimates will be inflated. This second problem is similar to ignoring the effects of blocking (a known source of variation).

To better understand these problems it is convenient to write down an explicit expression for the statistical model (e.g. JK):

$$y_{ijk} = \mu + s_{ik} + \pi_j + \tau_{d[i,j]} + e_{ijk}$$

where μ is the intercept, π_j is the period effect of period j =1,2, $\tau_{d[i,j]}$ is the direct treatment effect of the treatment given in period j of sequence i, s_{ik} is the random subject effect of subject k in sequence i, and e_{ijk} is the random error for subject k in period j in sequence i (for the moment we ignore carry-over effects). From that model, the fixed effects for each period and sequence for a 2x2 design are shown in Table 4.1 (p. 92).

Table 4.1: Fixed-effects for the 2x2 design. In this table, carry-over effects have not been included; including them would result in the fixed effects for period 2 being $\mu + \pi_2 + \tau_2 + \lambda_1$ and $\mu + \pi_2 + \tau_1 + \lambda_2$, in sequences AB and BA respectively.

Sequence group	Period 1	Period 2
AB	$\mu + \pi_1 + \tau_1$	$\mu + \pi_2 + \tau_2$
BA	$\mu + \pi_1 + \tau_2$	$\mu + \pi_2 + \tau_1$

The expected value of the difference A-B for animals from sequence AB (dAB_{AB}) is $(\tau_1 - \tau_2)+(\pi_1-\pi_2)$, and the expected value of the difference A-B for animals in sequence BA (dAB_{BA}) is $(\tau_1 - \tau_2) + (\pi_2 - \pi_1)$. The paired t-test is the same as testing if the set of all dAB_{AB} and dAB_{BA} are centered around zero, using a one-sample t-test. If there are more animals in AB than in BA, our estimate of treatment effects $(\tau_1 - \tau_2)$ will be biased by a factor proportional to $(\pi_1 - \pi_2)$; when the sample sizes of both sequences are the same, there will be no bias in the estimate of the treatment effect, but the error term will be inflated by a term proportional to $(\pi_1 - \pi_2)^2$. Thus, a paired test results in biased estimates of treatment effects and/or inflated variance estimates; counterbalancing, per se, does not result in a correct analysis, contrary to what is sometimes believed.

To prevent these problems, we should use the Hills-Armitage approach, illustrated in Table 4.2 (p. 93) and described in more detail in JK (p. 23-28), SN (p. 42-44), and Crowder & Hand (1990, p. 101). We take period differences (subtract period 2 from period 1) for both sequences, yielding $d12_{AB}$ and $d12_{BA}$ for animals from sequences AB and BA respectively. The expected values of these differences are: $E(d12_{AB}) = (\tau_1-\tau_2)+(\pi_1-\pi_2)$, $E(d12_{BA}) = (\tau_2-\tau_1)+(\pi_1-\pi_2)$. We can test for treatment differences comparing the means of $d12_{AB}$ and $d12_{BA}$ ($\overline{d12_{AB}}$ and $\overline{d12_{BA}}$) between the two sequences (e.g., a two-sample t-test). Define $\hat{\tau} = 0.5(\overline{d12_{AB}} - \overline{d12_{BA}})$; its expected value is $(\tau_1 - \tau_2)$ (so there is no bias) and the variance contains only a term for the within-individual errors (see expression in JK, p. 26). In other words, to test for treatment differences we compute the mean between the first and the second period for each individual, and then we use a two-sample t-test to compare these values between the two sequences. This

Table 4.2: Simulated data (columns three and four) for a 2x2 trial. A common (incorrect) analysis of treatment effects uses a paired t-test, which is the same as testing if the crossover differences are centered around zero. The Hills-Armitage approach compares period differences between the two sequence groups.

Sequence	Subject	Period 1	Period 2	Period differences	Crossover differences
				$d12_{AB}$	dAB_{AB}
AB	1	16.5	11.1	5.4	5.4
AB	2	14.9	9.2	5.7	5.7
AB	3	14.2	6.9	7.3	7.3
AB	4	20.6	13.8	6.8	6.8
AB	5	18.2	12.8	5.4	5.4
				$d12_{BA}$	dAB_{BA}
BA	6	15.0	13.3	1.7	-1.7
BA	7	13.9	9.8	4.1	-4.1
BA	8	9.8	6.5	3.3	-3.3
BA	9	16.8	14.8	2.0	-2.0
BA	10	14.9	12.0	2.9	-2.9

method of testing for treatment effects is also called the CROS test.

To test for period effects, we compute cross-over differences (difference between periods 1 and 2 for subjects in AB, and difference between periods 2 and 1 for subjects in BA — equivalent to computing differences between A and B for all subjects), and use a two-sample t-test comparing these differences between the two sequences. Finally, to test for inequality of carry-over effects we compare the sum of the values in the two periods between the two sequences (see JK, p. 24-25); note that we cannot test for absence of carry-over effects, only inequality or differential carry-over effects (see next section), and in the 2x2 designs differential carry-over is confounded with sequence effects. A nonparametric version of these tests was first described by Koch (1972) and is explained in JK (p. 51 and ff.) (but see Taulbee, 1982, for corrections of expressions(4) and (6) in Koch, 1972 and JK, p. 27 and 56).

As an example, Table 4.2 shows a set of data from an AB,BA trial (these are simulated data,

Table 4.3: ANOVA table for the analysis of the data in Table 4.2 using split-plot (parameterization as in JK, except no carry-over included).

Source	d.f.	SS	MS	F	p-value
Between-subjects	9	130			
Within-subjects stratum					
Period (adjusted for Treatment)	1	99.5	99.5	231.7	0.0001
Treatment (adjusted for Period)	1	13.8	13.8	32.1	0.0001
Within-subjects residuals	8	3.4	0.4		
Total	19	246.7			

from a model with main effects of period and treatment and normally distributed subject and random errors). Using the paired t-test approach to test for treatment differences we obtain t_9=1.098, $p = 0.3$. Using the Hills-Armitage approach we obtain t_8=5.666, p =0.0005 (with the Hills-Armitage approach we have one less d.f. as this is a two-sample t-test). In this example the paired t-test fails because there are period effects, whereas the Hills-Armitage approach has no problems with the period effects.

We can also analyse these data using a split-plot ANOVA (Table 4.3; see JK, p. 30-33). The first stratum is individual; the second stratum is within-individual and is used for the tests of interest (treatment effects). In this ANOVA, we use as explanatory or independent variables treatment and period, and test for treatment effects after having entered period in the model (and for period after entering treatment); these are called marginal tests. In this ANOVA we have adjusted for the effects of period by incorporating period into the model, and thus we obtain the exact same results as the Hills-Armitage approach ($F = 32.1 = 5.666^2 = t^2$). (However, an ANOVA that did not include period would yield the same incorrect results as the paired t-test).

The problem of the paired comparison is the same regardless of whether we use a t-test, a nonparametric test, or a randomization test. The cause of the problem is not the type of statistic but failure to account for the effect of period. Unless there is strong evidence to the

contrary, in most behavioural experiments we should assume that period can affect the results; in this case, a paired test should not be used because it is inappropriate, regardless of whether or not counterbalancing is used and whether or not there are the same number of subjects in each sequence. Problems with period effects are not limited to two-treatment cross-over designs, but affect all other designs as well (e.g., three treatment designs).

4.5 Carry-over effects

A potential problem of cross-over designs are period*treatment interactions (the effect of a treatment is not constant over the different periods). One type of period*treatment interaction is carry-over effect: the response to a treatment is affected by what treatment was applied in previous period(s), so that past treatments have effects that last, or carry-over, to the following periods. In the 2x2 design, but not necessarily in designs with more than two treatments or periods, carry-over and any other treatment*period interactions are completely confounded.

Carry-over effects can bias the estimates of treatment effects and affect designs with any number of periods and treatments. In most designs (including the 2x2), the cause of the problem is not carry-over per se, but differential carry-over effects, i.e., the carry-over from different treatments being different. For example, in Table 4.1 (p. 92), if there are differential carry-over effects, our estimate of treatment effects using the Hills-Armitage approach will be biased by $\lambda_1 - \lambda_2$; if there are equal carry-over they will be indistinguishable from period effects, and the Hills-Armitage approach will be unbiased. (Using a paired t-test, differential carry-over effects will result in bias, even if there are no period effects).

Contrary to what is sometimes believed, counterbalancing does not eliminate bias caused by carry-over effects, regardless of the number of treatments (e.g., Abeyasekera & Curnow, 1984). Thus, there are two strategies for dealing with carry-over effects: a) minimise the chances that they can happen; b) include them explicitly in the statistical model. Which one of these

approaches is taken will affect both the design of the experiment and the analysis of the data.

For the 2x2 design, there has been considerable debate on how to deal with carry-over effects. In the two-stage approach one first tests for carry-over effects and if no carry-over is detected one then tests for treatment effects with the CROS test (see p. 93); if carry-over is detected only the first period is used and one tests for treatment effects with the PAR test (as if we were dealing with a parallel groups design). The problem is that the results from the two-stage approach are either the same as for CROS or have an unknown but possibly very large bias, as the results from the PAR test and the test for carry-over effects are highly correlated (SN, p. 52-54; Grieve & Senn, 1998). This suggests that the two-stage approach should not be used. On the other hand, it is debatable if we can trust the results of the CROS test without first testing for carry-over (Jones & Wang, 1998). Tudor & Koch (1994; also Koch, 1998) have proposed a three-stage procedure; it is not known if this three stage procedure performs much better than the two-stage one.

With more than two periods, by making some assumptions it is possible to eliminate the problems from carry-over effects by including carry-over effects in the statistical model. For example, with 1st order carry-over effects, some designs (strongly-balanced designs; see Tables 4.4, p. 100 and 4.5, p. 101) result in estimators of treatment effects that are not affected by the presence of carry-over effects. However, the assumptions that allow us to include carry-over in the statistical model effectively might be unrealistic. One common assumption is the absence of second-order carry-over (i.e., effects that carry-over two periods after the treatment was applied); lack of second-order carry-over is frequently justified arguing that second-order carry-over effects are unlikely if there are no first-order carry-over effects. A second common assumption is the absence of carry-over*treatment interactions; carry-over by treatment interactions occur, for instance, when a treatment can carry-over into other treatments but it cannot carry-over into itself or when the effect of carry-over depends on the treatment into which it carries over. Depending on the underlying biological phenomena these might either be

reasonable approximations or completely inappropriate assumptions.

Senn (SN, ch. 10) discussed several reasons why models with carry-over effects are of no use, emphasising that many of the above assumptions are unrealistic. He shows that carry-over adjusted estimates can be even more biased than estimates unadjusted for carry-over. Thus, Senn (SN; e.g., p. 14-15; ch. 10) advocates using sufficiently long times in between application of treatments (wash-out periods) so that carry-over effects are very unlikely, and analysing the data without ever attempting to adjust for carry-over effects. The practitioner, however, should be aware that the results are conditional on the assumption of no carry-over effects. Moreover, in many studies (e.g., comparison of a control with an active treatment) if carry-over is present but not accounted for it will tend to underestimate the treatment difference (Jones & Lewis, 1995; SN, p. 102). In other words, carry-over will result in a decrease in power but not an increase in the probability of rejecting the null hypothesis when it is true (Type I error rate). In contrast to Senn's approach, there is a large statistical literature that models carry-over effects (e.g., JK, REA) and some authors strongly advocate always including carry-over effects (e.g., Abeyasekera & Curnow, 1984).

Unfortunately, in many behavioural ecology studies not enough information is available to determine what is a long enough wash-out period. A practical solution might be as follows: first, design studies so that carry-over effects are unlikely. The experimenter's attitude towards carry-over effects should be explicit. Second, design experiments so that carry-over effects can be included in the statistical model, (modelling carry-over effect in the most reasonable way). If carry-over turns out to be present, a design that made a provision for carry-over would make it possible to salvage the experiment, and would indicate that future experiments might need to increase the wash-out period.

Moreover, in some studies presence of carry-over effects after what was considered a sufficiently long wash-out period could reveal a phenomenon of interest in its own right, since a carry-over effect would indicate that a past experience is much longer lasting than expected (e.g.,

effects of prior defeats in aggressive encounters that affect fight performance more than 24 h after the defeat). Finally, in some instances we might combine cross-over designs with between-subject designs (e.g., Díaz-Uriarte & Marler, in prep.); an interaction between carry-over and between-subjects treatment might indicate a potentially interesting biological phenomenon. For instance, we might examine simultaneously the effect of hormonal treatment (a between-subject treatment) and effects of presentation of a female vs. a control (using a cross-over trial). In a study like this, an interaction between carry-over and hormone treatment would provide evidence that hormonal treatment has affected how long-lasting the presentation of a female is.

4.6 Design of cross-over trials

Here I discuss the main designs that could be useful in behavioural studies; more details are provided in JK, SN, and REA. I will only examine designs that consider period effects plausible. To maximise power, subjects should be allocated to treatments so that there are equal numbers of subjects for each sequence (and this restriction should be reflected when using randomization tests).

During the design phase, it is essential to understand how the data will be analysed. For example, some nonparametric methods for more than two treatments require that the designs be of a specific kind or that allocation of subjects be done in a particular way; some other methods only work with large sample sizes. These requirements might prompt one to either change the design, to try to allocate more subjects or allocate subjects in different ways, or to measure different response variables.

4.6.1 Designs for two-treatment trials.

The most common cross-over design is the AB,BA design. As we have seen, this design is problematic in the absence of information about carry-over effects. Even when carry-over effects

are not present designs with more than two periods can be preferable as they lead to estimators of treatment effects with smaller variance, and therefore increase power (e.g., the ABB,BAA design has a variance for the estimate of treatment effects which is 19% of that from AB,BA –provided we use the same number of subjects, allocated in equal numbers to each sequence).

Table 4.4 (p. 100) shows three two-treatment designs, and some of their basic properties which affect the degree of aliasing (aliasing refers to the presence, in the design matrix, of covariates which are linear combinations of other covariates; technically, it refers to the amount of overlap between the subspaces defined by the covariates; McCullagh & Nelder, 1989, p. 61-68). The consequence of aliasing is that we cannot obtain separate estimates of each parameter. Aliasing is a common problem in cross-over designs; the correlation between parameters is an indication of aliasing, and is listed for many designs in JK (and can also be obtained by matrix operations from the design matrix; see, e.g., REA). For instance, in the design ABB,BAA the correlation between the estimate of treatment and carry-over effect is zero, and thus the estimate of treatment effects is the same in a model with or without carry-over effects, which is a good quality if the statistical model includes carry-over effects.

We can classify two-treatment designs by the number of sequences and the number of periods. Designs differ in the variance of estimated treatment effects (tabulated in JK for many designs). In general, the more periods the smaller the variance, but when sequences with many periods are used it is more likely that there will be missing data for later periods; thus, designs with more than 3 or 4 periods are not very advisable. Also, some designs are less affected by having to end a trial before it was expected: if one uses a design such as ABBA,BAAB and cannot collect data from the last period one is left with ABB,BAA which is a good design (in contrast with eliminating the last period from AAAB,BBBA). When only two periods can be used the AA,BB,AB,BA design (Balaam's design for two treatments) can minimise problems from carry-over effects; however, this design might be a worse choice than simply ensuring a long enough wash-out period and using AB,BA.

Table 4.4: Some cross-over designs for two treatments(see JK; definitions from Vonesh & Chinchilli, 1997 are slightly different from those in Laska et al., 1983 and JK).

Design	Uniform within sequences[1]	Uniform within periods[2]	Balanced[3]	Strongly balanced[4]	Variance of the estimator of treatment effects[5]	Variance of the estimator of treatment effects when carry-over effects are present[5]
ABB,BAA	No	Yes	Yes	Yes	0.375	0.375
ABBA,BAAB	Yes	Yes	Yes	No	0.250	0.275
ABBA,BAAB, AABB,BBAA	Yes	Yes	Yes	Yes	0.250	0.250

[1]A design is uniform within sequences if each treatment appears the same number of times within each sequence;sequence effects are not aliased with treatment effects.
[2]A design is uniform within periods if each treatment appears the same number of times within each period; period effects are then not aliased with treatment effects.
[3]A design is balanced if each treatment precedes each other treatment the same number of times; in this case, carry-over effects are aliased with treatment effects. A balanced design, as defined in JK, is one that is balanced (as in this table), uniform within sequences –actually, each subject receives each treatment only once– and uniform within periods, and with equal number of subjects per sequence.
[4]A design is strongly balanced (or completely balanced) if each treatment precedes each other treatment, including itself, the same number of times; in this case, carry-over effects are not aliased with treatment effects.
[5]Expressed in multiples of (σ^2/Total number of subjects), assuming equal numbers of subjects allocated to each sequence.

Designs composed of many sequences will be more complicated to use, in particular with limited sample sizes, as one will need sample sizes which are integer multiples of the number of sequences (to have the largest power). This is more problematic when one uses blocking or between-subject treatments (as one usually will want to use the complete design –i.e., all the sequences– in each block or between-subject treatment). Designs composed of dual sequences (i.e., pairs of sequences where the second sequence is obtained by interchanging the treatment labels A and B of the first sequence) allow one to use simple and robust analysis based on within-individual comparisons (see JK, SN; also DU2). The designs in Table 4.4 (p. 100) are composed of dual sequences and are among the most useful for estimating treatment effects

Table 4.5: Examples of cross-over designs for four treatments; a) Williams design; b) for every pair of treatments two sequences can be found where the treatments appear in interchanged periods (e.g., in sequence 1, A is in the 1st period and D in the 2nd period, whereas in sequence 4 the positions of A and D are reversed.

a)

Sequence	Period			
	1	2	3	4
1	A	D	B	C
2	B	A	C	D
3	C	B	D	A
4	D	C	A	B

b)

Sequence	Period			
	1	2	3	4
1	A	D	B	C
2	B	C	A	D
3	C	B	D	A
4	D	A	C	B

and also perform well under different within-individual correlation structures (JK; Matthews, 1990).

4.6.2 Designs for more than two treatments

With more than two treatments we can distinguish between variance balanced (all pairwise differences between treatments are estimated with the same precision) and partially balanced designs (the variance of the comparison between two treatments depends on which two treatments are compared). Partially balanced designs might be the best choice when there are several experimental treatments and one control and we are most interested in minimising the variance of contrasts between each experimental treatment and the control. We can also differentiate between complete and incomplete block designs (e.g., SN, p. 163 and ff.; JK, p. 199 and ff.); in the latter the number of treatments is larger than the number of periods (so each individual is not subject to all the treatments). Incomplete block designs are particularly useful with large numbers of treatments; however, these are much more difficult to design and analyse, and thus are of limited interest in animal behaviour studies.

If period can have an effect (as we generally assume), designs should be uniform within periods (see Table 4.4, p. 100, for explanation). Designs uniform within periods can be based

on Latin squares (briefly, suppose we arrange our design as a square, with n rows and n columns; then, in a Latin square we can apply n treatments, and ensure that each treament is applied once, and only once, in each row and column; for cross-over designs, the rows represent sequences and the columns represent periods). Williams designs (e.g., Table 4.5a) are also balanced (with respect to carry-over; see Table 4.4). Under certain assumptions, we can minimise problems from carry-over effects by using extra-period designs. For example, we can use a Williams designs to which we add a period so that the last treatment is equal to the previous one (e.g., in Table 4.5, the first sequence would be ADBCC), and we obtain a strongly balanced design (see Table 4.4, p. 100). However Williams and strongly-balanced designs might not be particularly useful if carry-over is not an issue. Other designs based on Latin squares (e.g., Table 4.5b) have the property that, for every pair of treatments two sequences can be found where the treatments appear in interchanged periods (SN, p. 122 and 123); this property allows us to use some nonparametric and multivariate analyses (see DU2). Discussion of designs for three or more treatments can be found in SN (ch. 5, 9, 10), JK (ch. 5), and REA (ch. 5 and 6). In general, designs for more than four treatments will require sample sizes larger than those available in most behavioural studies.

The assignment of subjects to sequences (including blocking), and the election of the number of squares, are discussed in SN (p. 123 & 209-210) and JK (p. 196-197; 198-199). In a three treatment trial, we can either use one or the two Latin squares (if carry-over effects are included in the model, we will use the two sets of Latin squares). For four treatments, either several squares or a single one can be used; the latter is generally simpler and will be less affected by loss of subjects.

Finally, the optimality of the designs discussed above depends on assumptions that might be inappropriate in some cases (e.g., when we expects treatment*carry-over interaction). It is possible to construct optimal cross-over designs tailored to the particular assumptions of our model (see Donev, 1998; Jones & Donev, 1996), and also use a sequential approach to trial

design, so that assumptions can be incorporated as information becomes available.

4.6.3 Between-subjects designs and baseline data

Cross-over designs can be used in experiments that also include between-subject treatments (e.g., comparing the effect of female presence/absence in a cross-over trial, in which different individuals have been assigned to different hormonal manipulation treatments). Inclusion of these between-subject factors in the analyses is reviewed in DU-2.

The use of baseline data (data collected before treatment(s) is(are) applied) can be found in JK and SN (see also Tsai & Patel (1996) for non-parametric analysis of a 2x2 design). Baseline data can increase the sensitivity of tests for treatment*period interactions and between-subject treatments; however, baseline data do not increase sensitivity of tests of direct treatment effects, and thus are unlikely to be useful in most behavioural studies.

4.7 Conclusions

Cross-over designs can result in an increase in power and reduce the number of animals needed in a study, which is particularly important if there are ethical concerns or we are working with small and/or threatened populations. However, the analysis of cross-over trials tends to be more complicated than the analysis of parallel trials, and the potential for aliasing of effects in cross-over designs is larger; in addition, cross-over trials require that subjects be used repeatedly. Thus, election of cross-over designs vs. parallel trials will have to consider how costly it is to obtain new subjects vs. how costly it is to obtain repeated measures of the same subject. Additional (but rarely available) information on within- vs. among-individual variance would allow more informed choices between cross-over and parallel group designs (see details in SN, ch. 9).

In many studies conducted in the lab or in field enclosures that require lengthy training or habituation of animals, cross-over trials are probably good choices (if not the only option). In some field studies relocating subjects might be too time consuming compared to finding new subjects, whereas other field studies use individually-marked animals that can be relocated easily. However, even when subjects are easy to relocate, cross-over designs might be difficult to use in field conditions: the assignment of subjects to sequences will have been done before the animals are actually found on a particular day, and for period to have the same meaning across subjects, the time interval between periods should be comparable among animals. These conditions might impose too many constraints on which particular animals need to be found on a particular day, and could make cross-over designs less attractive.

The type of response will also affect the design of choice (see DU-2: next chapter). Thus, during the design stages (i.e., before any data have been gathered) it is very important to decide upon and understand the types of analyses that will be used; this might show that certain analyses are not possible and could prompt a change in the design. It is too risky to assume that any design and type of data can be analysed statistically.

In summary, this paper has argued that: a) a large number of designs is available for behavioural studies; designs composed of dual sequences are usually preferable, and even when dealing with two treatments we might not want to limit ourselves to the 2x2 design (see Table 4.4, p. 100); b) we will (virtually) always have to include period in our statistical analyses; c) we need to think about carry-over effects and what constitutes an appropriate wash-out period; how we are dealing with period and carry-over effects should be made explicit.

4.8 Acknowledgements

C. Lázaro-Perea provided advice, discussion, and comments on the ms. A. R. Ives, B. Jones, J. K. Lindsey, C. A. Marler, C. T. Snowdon and B. C. Trainor provided comments on the

ms. E. V. Nordheim suggested that this paper should be written. The book by B. Jones and M. G. Kenward is a delightful and amazing book that opened my eyes to cross-over trials (and taught me a lot of what I know about them). I wrote this manuscript using LyX, an open-source (and free) LaTeXbased document processor released under the GNU GPL license. Manuscript preparation partially supported by a John and Virginia Emlen Graduate Student Fellowship, Dept. Zoology, University of Wisconsin-Madison.

4.9 References

Abeyasekera, S. & Curnow, R. N. 1984. The desirability of adjusting for residual effects in a crossover design. Biometrics, 40, 1071-1078.

Bailey, N. T. J. 1995. Statistical Methods in Biology, 3rd Ed. Cambridge: Cambridge University Press.

Bart, J., Flinger, M. A., & Notz, W. I. 1998. Sampling and Statistical Methods for Behavioral Ecologists. Cambridge: Cambridge University Press.

Campbell, R. C. 1989. Statistics for Biologists, 3rd Ed. Cambridge: Cambridge University Press.

Crowder, M. J. & Hand, D. J. 1990. Analysis of Repeated Measures. New York: Chapman & Hall.

Donev, A. N. 1998. Crossover designs with correlated observations. Journal of Biopharmaceutical Statistics, 8, 249-62.

Edgington, E. S. 1995. Randomization Tests, 3rd Ed. New York: Marcel Dekker.

Grieve, A. and Senn, S. 1998. Estimating treatment effects in clinical crossover trials. Journal of Biopharmaceutical Statistics, 8, 191-233; discussion 235-47.

Jones, B. and Donev, A. N. 1996. Modelling and design of cross-over trials. Statistics in Medicine, 15, 1435-46.

Jones, B. & Kenward, M. G. 1989. Design and Analysis of Cross-Over Trials. New York: Chapman & Hall.

Jones, B. & Lewis, J. A. 1995. The case for cross-over trials in phase III. Statistics in Medicine, 14, 1025-1038.

Jones, B. & Wang, J. 1998. Comments on "Estimating treatment effects in clinical crossover trials". J. Biopharmaceutical Statistics, 8, 235-238.

Koch, G. G. 1972. The use of non-parametric methods in the statistical analysis of the two-period change-over design. Biometrics, 28, 577-584.

Koch, G. G. 1998. Comments on "Estimating treatment effects in clinical crossover trials". J. Biopharmaceutical Statistics, 8, 239-242.

Laska, E., Meisner, M., & Kushner, H. B. 1983. Optimal crossover designs in the presence of carryover effects. Biometrics, 39, 1087-1091.

Lehner, P. N. 1979. Handbook of Ethological Mehtods. New York: Garland Press.

Martin, P. & Bateson, P. 1993. Measuring Behaviour, 2nd Ed. Cambridge: Cambridge University Press.

Matthews, J. N. S. 1990. The analysis of data from crossover designs:the efficiency of Ordinary Least Squares. Biometrics, 46, 689-696.

McCullagh, P. & Nelder, J. A. 1989. Generalized Linear Models, 2nd Ed. New York: Chapman & Hall.

Schneider, B. 1983. Crossover designs and repeated measurements. Neuropsychobiology, 10, 49-55.

Senn, S. 1993. Cross-Over Trials in Clinical Research. New York: John Wiley & Sons.

Sokal, R. R. & Rohlf, F. J. 1995. Biometry, 3rd Ed. New York: W. H. Freeman.

Taulbee, J. D. 1982. A note on the use of nonparametric methods in the statistical analysis of the two-period changeover design. Biometrics, 38, 1053-1055.

Tsai, K. T. and Patel, H. L. 1996. Robust procedures for analysing a two-period cross-over design with baseline measurements. Statistics in Medicine, 15, 117-126.

Tudor, G. and Koch, G. G. 1994. Review of nonparametric methods for the analysis of crossover studies. Statistical Methods in Medical Research, 3, 345-81.

Vonesh, E. F. & Chinchilli, V. M. 1997. Linear and Nonlinear Models for the Analysis of Repeated Measurements. New York: Marcel Dekker.

Zar, J. H. 1996. Biostatistical Analysis, 3rd Ed. New Jersey: Prentice Hall.

Zolman, J. F. 1993. Biostatistics. New York: Oxford University Press.

Chapter 5

Cross-over trials in animal behaviour. II: Analysis and plotting

5.1 Abstract

Cross-over trials are frequently used in animal behaviour experiments but are often analysed incorrectly (see previous chapter). In this paper I review methods of analysis of cross-over trials in the context of animal behaviour experiments. I group methods of analysis according to the type of response variable: non-parametric and robust methods for metric responses, parametric methods for metric responses —linear mixed-effects models—, models for categorical responses both non-parametric and parametric —extensions of generalized linear models—, censored observations —survival analysis–, and multivariate responses. Within-individual contrasts are explained in detail early on, as they are the basis of many different methods, from non-parametric to multivariate and survival-based models, and they offer a useful framework for extending the analysis of data from cross-over trials to situations where robust methods might be needed (e.g., permutation tests of censored multivariate responses). In this chapter I also discuss some types of plot that are specific and particularly useful for cross-over trials. Before conducting a study, it is of paramount importance to consider both the design and analysis, since the type of response can affect the choice of design. Moreover, some types of responses can be very difficult to analyse, specially with small sample sizes, and can result in very low statistical power (in particular categorical and survival data), and might prompt us to redesign

the experiment or consider measuring other responses.

5.2 Introduction

Cross-over trials are frequently used in animal behaviour (see Díaz-Uriarte, in review —previous chapter; hereafter DU-1) as they allow us to conduct experiments with relatively small numbers of subjects that nonetheless achieve high statistical power by using each subject as its own control (Jones & Kenward, 1989 —hereafter JK; Senn, 1993a —hereafter SN). Thus, cross-over designs are powerful tools when repeated testing of one subject is much simpler than recruitment of new subjects. However, cross-over experiments in animal behaviour studies are usually analysed incorrectly, as if they were matched pairs or "typical" repeated-measures designs, which they are not (see DU-1 for details and examples). The main problems are failure to account for period and carry-over effects. The widespread used of inappropriate analyses could be the result of a lack of information about cross-over trials in statistical texts commonly used by behaviourists. The problem is compounded because in many behavioural experiments researchers often record data (such as categorical data or censored time to event data) that might not allow the use of standard parametric analyses, and frequently measure several response variables that ought to be analysed with multivariate techniques.

The main objective of this paper is to review the analysis of cross-over designs in the context of animal behaviour experiments. This paper should be of immediate and practical use for behaviourists and statistical consultants working with behaviourists. I review and show the connections among different methods that have recently appeared in the statistical literature and are relevant to behaviourists (e.g., multivariate responses and time to event data), but that are not covered in available texts (JK; SN; Ratkowsky et al, 1993). On other topics (e.g., linear mixed-effects models) I provide practical discussion in the context of cross-over trials. Nonparametric and categorical data methods are considered in recent reviews of cross-over trials; I have included some new papers and eased the use of these methods by cross-referencing

statistics textbooks and software packages. Small sample sizes, blocking, and among-subject treatments are all relevant to animal behaviour experiments and are considered throughout the paper. I concentrate on methods that are available in major statistical packages (specially SAS, S-Plus, and R; note that R is free GNU software that can be obtained from CRAN at http://cran.r-project.org and mirror sites; unless specified otherwise, S-Plus libraries are available from Statlib at http://lib.stat.cmu.edu and R libraries from CRAN), or that can be implemented with a minimum amount of code writing. Finally, I emphasise randomization and permutation tests (e.g., Edgington, 1995; Good, 1994; Noreen, 1989). Randomization tests, increasingly used in behaviour and ecological research (e.g., Manly, 1997; Crowley, 1992), are a general alternative when parametric assumptions are not met, can be more powerful and flexible than traditional "non-parametric" methods, and might be the most appropriate tests for many experimental settings (Ludbrook & Dudley, 1998).

I review the analysis of data from cross-over designs according to the type of response variable (e.g., Agresti, 1990, ch. 1). A metric or interval response is one that has numerical distances between any two levels of the scale (e.g., length); arithmetic operations on the response are meaningful. One special type of metric responses is time to an event (examined later). Ordinal variables are categorical variables that have ordered levels (e.g., bad, fair, good), but differences, sums, and other algebraic operations on the ranks or levels are not meaningful. Nominal categorical variables have levels without natural orderings (e.g., Buddhist, Christian, Hindu). A particular type of categorical responses are binary outcomes (such as success/failure).

I first review the analysis of data from cross-over experiments. Next I cover plotting and graphical summaries in cross-over experiments. Then I discuss sample size and missing data. I conclude with some recommendations on the use and analysis of cross-over experiments in animal behaviour experiments. Elsewhere (DU-1 —previous chapter) I have reviewed some basic terminology and the design of cross-over trials.

5.3 Metric responses: nonparametric and robust methods

5.3.1 Within-individual contrasts

With two treatments and dual designs, a common way to carry out robust analyses (JK, p. 60-65; 160; Hafner et al., 1988) is to use within-individual linear contrasts to reduce the data from each individual to a single number and then compare these numbers between sequences. The use of within-individual contrasts is the basis of many analyses of cross-over trials (including some multivariate analyses), and thus will be explained in detail.

The within-individual contrasts are linear functions of the observations of each subject; the contrasts' coefficients are the same for all sequences, and the sum of the contrasts' coefficients adds to zero. The estimator of the effect of interest is the difference between (the mean of the) within-individual contrasts of the two sequences. For example, in the 2x2 design the contrast for treatment effects is the difference between the measures in the first and second periods; we obtain the estimator of treatment effects as the difference between the mean contrasts from sequences 1 and 2 (see JK —pp. 23-28— and SN —pp. 42-44; also DU-1).

Contrasts are chosen so that they isolate the effects we are interested in (e.g., treatment effects). In the 2x2 design the Hills-Armitage analysis (explained in DU-1) is an example of the within-individual contrasts logic. In more complicated designs, there can be several possible linear contrasts for a particular effect, but the estimators with the smallest variance are the Ordinary Least Squares (OLS) estimators (Hafner et al., 1988). The design matrix $\mathbf{X}$ (with one row per cell mean) used to obtain the OLS estimators includes subject, treatment, and period effects and, if appropriate, carry-over effects (see Ratkowsky et al., 1993, for examples with cross-over designs). SN (p. 238-248) shows how to obtain the estimators without using matrices. Although these OLS estimators are, strictly, only optimal for uniform covariance structures, with other covariance structures the estimators are less efficient but are still unbiased (see JK) and will not result in increased Type I error rates. These estimators will all take the

form of a difference between groups of contrasts among the periods and all the valid contrasts must have the same form in the two sequences.

Once we obtain the contrasts, we compare them between the two sequences. As a test statistic we can use the difference between the two sequences of the mean (within sequence) of the within-individual contrasts. The p-value for this test can be obtained from a randomization test (e.g., Edgington, 1995), an independent t-test, or a Wilcoxon rank sum (=Mann-Whitney) test (e.g., JK, p. 51-60; SN, p. 93); Tudor & Koch (1994; hereafter TK) use the quadratic statistic given by their eq. 2.8 instead of a t-test. If using a Wilcoxon rank sum test, the ranking is done after the linear contrasts are applied (i.e., we do not rank the original data). Covariates (if their value remains the same over all periods of an individual) and other factors can also be examined by using as a response variable the within-individual contrasts (instead of the original values themselves) in a linear model that includes the covariates (e.g., Hafner et al., 1988). If using randomization tests, the restrictions in the allocation of subjects (e.g., same number of subjects to each sequence) should be taken into account.

In some designs (e.g., ABBA,BAAB), the variance of the estimator of treatment effects is smaller when no carry-over effects are included in the model. We can start with an OLS estimator from a design matrix that includes carry-over effects, and if the test of carry-over effects clearly indicates that these effects are unlikely, we could obtain a new OLS estimator of treatment effects from a design matrix that includes no carry-over effect (e.g., Hafner et al., 1988).

An analysis based on within-individual contrasts is robust in the sense that it makes no assumptions about the covariance structure (JK, p. 65, 160, 283; Hafner et al., 1988), although the analysis does assume that the responses of different animals are independent. However, in particular in designs with many periods, power is lost with respect to, say, a linear mixed-effects model when assumptions of the mixed-model are met.

The use of contrasts can be understood in a randomization test context. Under the null

hypothesis of no treatment effects, an individual that was assigned to sequence AB would have yielded the same pair of values if it had been assigned to sequence BA, because individuals are assigned randomly to sequences. Thus, the difference between periods 1 and 2 should be the same regardless of sequence assignment (note that possible period effects are thus taken into account). Contrasts must be the same regardless of sequence: under the null hypothesis, a linear combination of an individual's responses must remain unchanged —i.e., the estimate must be invariant under permutations of the observations. For most designs we cannot test for period effects using a randomization test: as periods are not randomized, the order of observations must remain the same in all possible random assignments of subjects to sequences (Shen & Quade, 1983). In fact, the OLS estimator for period will differ depending on the sequence, and we cannot devise a randomization test to examine period effects.

Transformations of data can affect the results of nonparametric and randomization methods. Before conducting any analyses, we should consider the appropriate scale for the data; e.g., if the effect of treatment will be to increase the response in one treatment by a multiple of the response under the other treatment (i.e., a multiplicative effect) then we will probably want to log-transform the data before any tests. Notice, however, that interpretation of results from parametric and non-parametric tests can differ (e.g., Conover, 1980; Johnson, 1995; Stewart-Oaten, 1995; Seaman & Jaeger, 1990).

5.3.2 Blocking, among-subject treatments, and more than two sequences

When experiments are carried out in blocks (e.g., weeks, age groups, or locations), analyses that use randomization tests can be applied as before, but the randomization tests must preserve the restricted randomization used in the experiment (e.g., Edgington, 1995; p. 131; Noreen, 1989, p. 28; Maritz, 1995, p. 191). The test statistic is computed from all data together for each permutation, but the random reallocation is restricted to within-blocks. Designs that involve both among and within-individual level treatments can be analysed with the approach

above, although care is required in the selection of the test statistic and the specification of the underlying model (e.g., interactions between the among and within-individual treatments should generally be considered). An alternative is to use the extended Mantel-Haenszel test (e.g., Agresti, 1990, p. 283, Koch & Edwards, 1988, p. 418; for cross-over TK, p. 358 and 375; there are several tests which contain the words "Mantel-Haenszel"; the test referred to here is applicable to ordinal response variables). With small sample sizes, this statistic's approximate chi-square distribution (1 d.f.) is not appropriate, and the p-value should be determined with a randomization test.

Designs made by pairs of dual sequences can be analysed like a blocked design, but now each sub-design will have its corresponding OLS estimator. The analyses using t-tests involve obtaining a combined estimator of the treatment difference and its variance, and are shown in JK (p. 171 and ff.). Alternatively, with randomization tests, the testing procedure would be analogous to a blocked design, where each sub-design constitutes a block (e.g., TK, p. 376); however, in contrast to the blocked design, here the test statistic is computed separately for each of the designs, and later combined (after weighting by sample size of each sub-design). For the AA,BB,AB,BA design see Elswick & Uthoff (1989; also TK, p. 374).

5.3.3 More than two treatments

Non-parametric tests of designs for three or more treatments are more complicated. SN (p. 144-152) presents a test that can be applied to designs with the appropriate structure (e.g., previous chapter, Table 4.5b, p. 101); the procedure is analogous to the one used for designs made of dual sequences (see paragraph above), where we test differences between pairs of treatments by arranging sequences in pairs where the two treatments appear in interchanged periods (analogous to dual designs). For each pair, we obtain the statistic by forming the appropriate within-individual contrasts. We then combine the statistics over all pairs of sequences using a weighted sum. This is another example of the extended Mantel-Haenszel test, and can be anal-

ysed as such (SN, p. 150; Koch & Edwards, 1988). Application of this test requires a particular (and somewhat restrictive) design; if we suspect we will use nonparametric methods, we should design the trial to conform to this structure in advance.

For designs that do not have this structure, Peace & Koch (1993) present a more general test, which is based on obtaining sequence differences of period contrasts, so as to isolate the effects of interest (e.g., pairwise differences between treatments). This method requires relatively large sample sizes and that the different sequences have the same number of subjects; allocation of subjects to sequences during the execution of the experiment should be done by blocks (with number of subjects per block an integer multiple of the number of sequences). A randomization test for a three-period, three-treatment trial is shown in Shen & Quade (1983); it can handle missing data, but assumes uncorrelated errors.

Tests for three-treatment, three-period designs that consist of replicated sets of two Williams squares are shown in Bellavance & Tardiff (1995). These tests are based on a non-parametric test of a randomized block design (a procedure similar to, but more efficient than, Friedman's test); it assumes that correlation of errors across time does not change, and it can not be extended to more than three treatments. For the s-treatment, s-period ($s \geq 3$) Williams square design, Ohrvick (1998) presents tests for treatment effects (and procedures for multiple comparisons); these tests also assume that correlation of errors across time does not change.

5.4 Metric responses: linear mixed-effects models

The distinguishing features of cross-over designs (e.g., JK; Lindsey, 1993) are time-changing covariates (the most obvious one being the within-individual treatment; other within-individual covariates might also change over time) and potentially correlated observations within individuals. Covariates can easily be considered in linear mixed-effects models, and these models can also be used to analyse complex experimental designs. Traditionally, cross-overs (and other re-

peated measures designs) were analysed with split-plot ANOVA. With more than two periods, however, the split-plot analysis makes restrictive and potentially unrealistic assumptions about the covariance structure (the so-called sphericity condition that, for example, implies that differences between responses in any two periods have the same variance). There are ways to deal with these restrictive assumptions (e.g., Diggle et al., 1994; Crowder & Hand, 1990), but it is generally more satisfactory to directly model the covariance structure using linear mixed-effects models (see Pinheiro & Bates, 2000; Littell et al., 1996; also Verbeke & Molenberghs, 1997; Bennington & Thayne, 1994; Lindsey, 1993). Mixed models are ideally suited for cross-over experiments as the latter include both fixed effects (treatment, period, carry-over) and random effects (the subjects or animals). Moreover, software for linear mixed models allows flexible modelling of the covariance structure, deal much better with unbalanced data than traditional ANOVA, and allow use of covariates that change both at the within and among-individual level. Additionally, mixed models can recover information about treatment effects available between subjects (Littell et al., 1996), which can be important in cross-over designs with unbalance (Brown & Kempton, 1994), either from missing data or by design —e.g., partially balanced designs. Finally, linear mixed-models are natural for examining questions of repeatability and individual differences (an important topic in animal behaviour —e.g., DeWitt et al., 1999; Aragaki & Meffert, 1998; and references therein), as they make it possible to test the relevance of the among-individual variance component.

Linear mixed models can be fitted using, for example, S-Plus and R (library nlme), SAS (PROC MIXED), as well as BMDP, Genstat, and others. Examples with cross-over trials are presented in Vonesh & Chinchilli (1997, ch. 4), Littell et al. (1996, pp. 392 & ff.), Lindsey (1993; pp. 136 & ff.). Aside from the modelling of covariance structure and variance heterogeneity, mixed models have many similarities with the usual linear models. An overview of the theory of linear mixed models can be found in Pinheiro & Bates (2000) and Littell et al. (1996) (see also Davidian & Giltinan, 1995, ch. 3). General strategies for model building are discussed in Pinheiro & Bates (2000) and Diggle et al. (1994; specially ch. 4 and 5) (see also Verbeke

& Molenberghs, 1997); in the context of cross-over designs, see Vonesh & Chinchilli (1997, ch. 4). Diagnostic plots of fitted models are covered in detail in Pinheiro & Bates (2000; see also Verbeke & Molenberghs, 1997). Mixed models present some difficulties with selecting the appropriate degrees of freedom to use when testing fixed effects (Brown & Kempton, 1994 —but with large F-values the differences in d.f. are inconsequential), and can be questionable with small sample sizes (in particular for the effect on estimation of the covariance matrix).

Because of the problems with carry-over effects, there has been disagreement about the appropriate parameterisation of the 2x2 design (e.g., see Ratkowsky et al., 1993, ch. 3). One parameterisation, based on JK (p. 30) is

$$y_{ijk} = \mu + \lambda_i + s_{ij} + \text{other.random} \ + \pi_k + \tau_{d[i,k]} + \text{other.fixed} + e_{ijk}$$

where in the fixed effects part μ is the intercept, λ is the carry-over (which in this parameterisation is equivalent to a sequence effect), π is the effect of period k, τ is the direct treatment effect of the treatment given in period k of sequence group i, s are independent and identically distributed (i.i.d.) N(0,σ_s^2) are the random effects of individual j in sequence i, and e i.i.d. N(0,σ^2) are the within individual errors. All random effects are independent of each other. "Other.fixed" refers to other fixed effects (covariates like body weight or temperature), and "other.random" refers to other random effects (e.g., blocks). A problematic aspect of this parameterisation for the 2x2 design is the inclusion of the carry-over effects (see discussion above).

A parameterisation that can be extended to models with more than two periods is

$$y_{ijk} = \mu + \xi_i + \text{other.fixed} + \text{other.random} + s_{ij} + \pi_k + \tau_{d[i,k]} + \lambda_{d[i,k-1]} + e_{ijk}$$

where everything is as above, but we have added ξ as the effect of sequence. e i.i.d. N($\mathbf{0}$,$\mathbf{R}$) is the random error associated with the m-th period measurement of subject k from sequence i, where $\mathbf{R}$ is the within-individual covariance matrix and is the same across levels of i, j, k. All

random effects are independent of each other. Here we can include both sequence and carry over effects. When there are more than two periods, the covariance structure should always be modelled appropriately. I have included in the later model a sequence effect; this is not done by JK or SN, but it appears in Vonesh & Chinchilli (1997, ch. 4; see also Lindsey, 1993, p. 15 and 135). We will generally want to include a term for sequence for three reasons. First, when fitting mixed models it is convenient to start with a "saturated model" to estimate the covariance structure (Diggle et al., 1994, ch. 4). Second, the sequence effect, if significant, might alert us to potential problems with the model; a significant sequence effect might result from bad luck during the randomization of subjects to sequences, but it could also be the result of higher order treatment*period and treatment*carry-over interactions not included in the model (see also Elswick & Uthoff, 1989). Third, in some cases sequence effects might be what are affected by among-subject treatments (i.e., we will find significant sequence by among-subject treatment interactions).

When modelling period effects it might be appropriate to initially model them as a categorical variable (as the effect of period might plateau), but it might be possible to obtain a simpler model by using polynomial contrasts and sequentially eliminating the higher-order terms, which could result in a model with just a linear trend with time. Moreover, modelling period as a continuous variable eliminates the confounding of period with carry-over (Ratkowsky et al., 1993). Finally, although a typical strategy of model building is generally employed (JK, but see SN), where non-significant terms are dropped from the model, the correct approach with non-significant carry-over effects is debated (e.g., JK, p. 150).

There are some differences in the literature on how to code the carry-over term. For example, suppose that our design has treatments A, B, C; we will need a carry-over column in our data with levels A, B, C, and 0 (Crowder & Hand, 1990, p. 107), as the first period has no previous treatment (but this means that the first period and carry-over 0 are completely confounded). This is the approach used by SN and Littell et al., 1996 (p. 392). However, in SAS we will

not be able to obtain estimates (e.g., LSMEANS statement); thus, Littell et al. (1996) recode carry-over, creating one dummy variable per treatment which has a 1 if that treatment was in the previous period, and 0 otherwise; this has no effect on the p-values, but allows to obtain estimates. We can also use dummy variables for both period and carry-over that avoid over-parameterisations (see e.g., Diggle et al., 1994, p. 156). In the example of three-periods and three-treatments, for period we use two dummies (say, x1 and x2), which take value 0 on the first period, and for carry-over we use also two dummies (say, x3 and x4), which take value 0 for previous treatment A; note that we do not need to code for the no-carry-over of the first period, as this corresponds to x1=0 and x2=0. This third coding strategy should produce similar results as the first two. The first two approaches do not work with nlme (S-Plus and R) if period is coded as a categorical variable, as we end up with a singular design matrix; however, the third will work in both SAS and S-Plus and R. Littell et al. (1991, p. 206) use a different method, which can yield different results from the above one. Ratkowsky et al. (1993) propose making the first carry-over (0) equal to one of the other treatments; this, however, is not recommended as results from mixed models depend on which other treatment is placed as the carry-over in the first period.

The d.f. that our analyses will yield should be examined during the design period, and also serve as a check of the software output (but beware that Satterthwaite's approximation might yield different d.f. in unbalanced designs). Following JK (p. 141), for a design with s sequences and p periods we will have (sp-1) d.f. that can be divided in (s-1) d.f. between groups, (p-1) between periods, and (s-1)(p-1) for the group*period effects (more will be available if period is modelled as a continuous variable). The latter (group*period d.f.) are the d.f. which relate to the effects of interest, specially treatment effects, treatment*period interactions, and carry-over effects. We can partition these d.f. in several different ways, but we will always be limited by the total (s-1)(p-1) d.f. (or more if period is continuous). JK discuss how some terms (in particular carry-over and treatment*period) might be aliased, which can affect the interpretation of treatment effects (see also Koch et al., 1983). With among-subject treatments,

some of the d.f. will be used to account for interactions such as treatment*among-subject treatment, period*among-subject, etc.

5.5 Categorical data

Categorical data are among the most difficult to analyse in cross-over designs; at the same time this is an area of very active statistical research. I start discussing several nonparametric-like methods, first for binary responses and next for ordinal outcomes. Later I review methods that are explicitly model-based.

5.5.1 "Nonparametric-like methods"

For the 2x2 trial with binary response, there are two main tests for treatment effects (see JK, p. 89-105; SN, p. 106-109; Crowder & Hand, 1990, p. 109-110; Fidler, 1984), and (as usual) these tests are appropriate for treatment effects in the absence of differential carry-over effects. Both tests are based on comparing scores for individuals in the two periods; each subject yields a pair of responses, cd, which means response c in period 1 and response d in period 2; thus, we can have pairs 00, 11, 01, 10 (the last two outcomes are referred to as showing a preference). The Mainland-Gart test uses only information form the 10 and 01 outcomes, comparing the number of each of these outcomes between the two sequences using, for example, Fisher's exact test. Prescott's test is equivalent to scoring profile 01 as -1, profile 10 as +1, and profiles 00 and 11 as 0, and comparing the mean profile between the two sequences using a randomization t-test (which is equivalent to using an exact conditional test for linear trend on the 2 x 3 contingency table —this is different from an exact test for independence). If the software package reports one-sided p-values for exact conditional tests for contingency tables we will want to double that p-value. The Mainland-Gart test does not depend on the random allocation of subjects to sequences, whereas Prescott's test does, but in virtually all behavioural ecology experiments

subjects will have been allocated to sequences randomly. Moreover, Prescott's test is generally more sensitive than the Mainland-Gart test. Thus, Prescott's test is likely to be the more useful of the two. However, tests of binary response data in 2x2 trials tend not be very powerful (i.e., they are not very sensitive to treatment differences), and this can be aggravated if only a few subjects in each sequence show a preference (i.e., are either 01 or 10). Becker & Balagtas (1993) present a test that can can be slightly more powerful than Prescott's test, but is also more complicated.

For binary responses and designs with three or more treatments and a particular structure (e.g., Table 4.5 b, p. 101), SN (p. 153-155) proposes a method analogous to the one described above for non-parametric analyses of metric responses with more than two treatments; this method can be applied with both Mainland-Gart's and Prescott's tests.

With ordered categorical data, Senn (SN, p. 109-113; for a detailed example see also Senn, 1993 b, and discussion by Ezzet & Whitehead, 1991, 1993) presents a simple method based on a heuristic argument for the 2x2 design. For each subject, we reduce the data from the two periods to another ordered categorical response (e.g., if in period 1 an individual was in good condition whereas in period 2 it was in very good condition, the value for this individual becomes "improve"). We are left with ordinal data for each sequence, and differences between the two sequence groups are an indication of treatment effects. We can compare the two sequence groups using, e.g., proportional odd models (Agresti, 1990, pp. 323-331). These methods might be questionable in trials with small sample sizes.

Alternatively, for ordinal data, TK (pp. 359-361) present several tests based on Wilcoxon's rank sum statistic; these tests involve differences between ranks within periods (in contrast to the other non-parametric tests where ranking was done over the whole sample). These statistics are easy to compute; with small samples, the p-value can be obtained from the permutation distribution. A more complicated approach is presented in Brunner & Newmann (1987) who use different tests based on alternative schemes of ranking the observations.

For 2-treatment, 2-sequences (and $\geq$2 periods) designs, Jung & Koch (1999) present a development of methods discussed in TK (p. 361-362) based on Mann-Whitney measures of association. In each period, these statistics estimate the probability of a larger response of a randomly selected member from one of the groups relative to a randomly selected member of the other group. This method allows stratification and inclusion of covariates and only requires moderate sample sizes ($\geq$ 10 individuals per sequence); the method is slightly complicated to apply (although Jung & Koch, 1999, present three detailed examples of application), but is useful for ordinal response variables and continuous asymmetric distributions (with possible outliers). Nonparametric methods for ordinal data with three or more treatments are not well developed.

5.5.2 Explicitly model-based methods

The methods in the previous section are specific for certain types of responses and/or designs. However, it is possible to analyse categorical data (binary, nominal, and ordinal) for a potentially unlimited range of cross-over designs with methods based on explicit models (see Kenward & Jones, 1994). These methods are based on generalized linear models (McCullagh & Nelder, 1989; Agresti, 1990; Dobson, 1990; Crawley, 1993). Generalized linear models are extensions of linear models that make it possible to analyse data in which a function —called the link function— of the mean response (but not the response itself) is linearly related to a set of predictors, and where the variance of the response might be a function of the mean response; generalized linear models have become the standard way of analysing categorical data.

With categorical data (and also with other data, such as survival; see below) we need to distinguish between different types of models, the two most common being marginal or population averaged, and subject-specific or random-effects (see discussion in Kenward & Jones, 1994; Albert, 1999; Diggle et al., 1994, ch. 7; Lindsey, 1993, ch. 2; Liang et al., 1992; Zeger et al., 1988). Briefly, marginal models model the marginal distribution of the response as

a function of the explanatory variables; this modelling is done separately from the within-subject correlation across time (which is treated as a nuisance) and the estimated coefficients have a population interpretation (not an individual interpretation). In contrast, in subject-specific models a random effect for an individual is introduced (as was done in the linear mixed models), and the parameter estimates (say, for treatment effects) modify the probability of a specific subject giving one response instead of another. The distinction between marginal and subject-specific models is not important for linear models because we can formulate the two approaches so that the coefficients have the same interpretation; however, with categorical (and survival) data this is generally not the case for most link functions.

Generalized estimating equations (GEE) are marginal models and can be implemented (see Horton & Lipsitz, 1999) using SAS (PROC GENMOD) and S-Plus and R (library gee; for S-Plus also library yags at http://www.biostat.harvard.edu/~carey); GEE's should perform relatively well in experiments with at least 20 subjects; estimators (e.g., of treatment effects) are consistent even when the correlation structure is misspecified, and testing is done using a robust estimator of variance; Albert (1999) and Horton & Lipsitz (1999) present useful tutorials on GEE's. However, J. K. Lindsey, has pointed out —pers. comm.— that GEE's are not appropriate for cross-over designs, because GEE's treat dependence among observations as if treatments were between subjects, instead of within subjects; thus, the corrected standard errors from GEE's are inflated instead of reduced —the opposite of what one wants—, and therefore result in lower statistical power. Generalized linear mixed models are subject-specific models in which the random subject effects are assumed to follow some distribution; these models can be fitted with SAS (PROC NLMIXED and macro GLIMMIX —Littell et al., 1996), and R (library repeated, from J. Lindsey, available at http://www.luc.ac.be/~jlindsey/rcode.html; see also Lindsey's libraries gnlmm for generalized non-linear mixed models and library growth) but might not perform adequately with small sample sizes. Conditional likelihood models are also subject-specific models (but here the subject effects are eliminated), and they can be fitted using software for log-linear models, such as SAS's PROC CATMOD (see Kenward & Jones,

1991, for examples), and for some conditional models distribution-free and exact permutation tests are available (Agresti, 1993; Kenward & Jones, 1994). Discussion and references of GEE's and generalized linear mixed effects models can be found in Albert (1999), Horton & Lipsitz (1999), Littell et al. (1996, ch. 11), Vonesh & Chinchilli (1997, ch. 8), Diggle et al. (1994, ch. 7-9), Kenward & Jones (1994), Lindsey (1993, ch. 2), Lipsitz et al. (1994), and SAS's on line manual (which includes a cross-over example). Recent examples of applications to cross-over trials are shown in Diggle et al. (1994; GEE's in pp. 154-159; conditional likelihood in pp. 175-181), Kenward & Jones (1994) and Lindsey (1993, pp. 201-204).

5.6 Time to event data: censored observations

Many studies in animal behaviour collect time to event data (also called failure time data or survival data) such as time until a certain behaviour is displayed (e.g., time to reemerge from a refuge following a predator's attack). Generally, animals are observed for a predetermined time, and the observer records when the event takes place. If the event takes place in every period for every subject, these are metric data (and can be analysed with either parametric or nonparametric methods). However, for some subjects the event might not occur within the observation period, which results in censoring (i.e., all we know is that the time till the event occurs is larger than the observation time). Although a small number of censored observations probably does not preclude the use of the parametric and nonparametric methods above, censored observations make usual techniques for metric data, including non-parametric ones (see France et al., 1991; Ducroq, 1997), inappropriate. Censoring can violate several of the assumptions of both parametric and non-parametric tests and will result in tests insensitive to treatment effects and biased estimates of treatment effects. In particular, converting survival data into 0/1 data (for no-event and event respectively) is not only arbitrary (the coding depends on the time at which the categorisation is made) but is also a very inefficient use of information. Moreover, 0/1 scores do not really facilitate the analysis with cross-over designs.

Censoring can be of several types (for details see, e.g., Klein & Moeschberger, 1997; Lee, 1992). The most common in behavioural studies is Type I censoring, where the event is observed only if it occurs before some predetermined time. This censoring time is usually common for all individuals; with random censoring —censoring time a random variable— data can be analysed with methods for Type I censoring, provided that censoring and survival times are independent (O'Brien & Fleming, 1987; Heimann & Neuhaus, 1998).

Analysis of censored data, generally referred to as survival analysis or reliability analysis, is well developed (e.g., Klein & Moeschberger, 1997; Collett, 1994; Lee, 1992; Lawless, 1982; Kalbfleisch & Prentice, 1980), but techniques applicable to experiments where the same individual experiences the event repeatedly are not common. Some methods have been proposed to analyse paired censored data (e.g., Woolson & O'Gorman, 1992; O'Brien & Fleming, 1987), but these methods cannot be applied to cross-over designs if there are period effects.

Two recent techniques available to analyse repeated time to event derive from the analysis of multivariate time to event data, but might not be appropriate with small sample sizes. The method developed by Lee et al. (1992; see also Lin, 1994, 1993; Wei et al., 1989) assumes a marginal proportional hazards model; it does not require that we specify the form of the joint distribution of the observations of each subject. Frailty models (e.g., Klein & Moeschberger, 1997, ch. 13; Ducrocq, 1997; Therneau & Grambsch, 2000) are subject-specific models in which all the observations from a subject share a common frailty (a common random effect that affects the hazard rates of all the observations of a subject); frailty models require that we assume a particular distribution for the frailty (generally a gamma). Both the marginal and frailty models are available in S-Plus and R (library survival5) and in SAS (PROC PHREG —Allison, 1995, pp. 236-247).

Lindsey et al. (1996) present a method specific for cross-over designs based on log-linear models, which has the advantage that it works with relatively small sample sizes and can be fitted with software that handles generalized linear models such as S-Plus, R, SAS, GLIM. The

R library event (available at http://www.luc.ac.be/~jlindsey/rcode.html; the syntax for model building with this library is somewhat different from other R statistical models) will fit these (see function ehr) and other models for repeated censored data. Segal & Neuhaus (1993) present a related marginal method that combines Poisson regression with GEE and can be implemented with SAS, S-Plus, or R. Two advantages of all these four methods are: a) they can accommodate covariates and factorial designs that mix within- and among-subject treatments —although not necessarily nested designs; b) they can be used to analyse experiments where we have measured more than one response variable. Many modelling strategies for these methods are common with linear models (see above).

Feingold and Gillespie (1996) suggested two nonparametric-like approaches for two-treatment designs. Their second method is tailored to the 2x2 design but is difficult to extend to other designs. Their first method has wider generality; one first ranks (see below) all the observations, and then applies the procedures for complete data to these ranks (i.e., one applies within-individual contrasts to the ranks, and later compares the within-subject contrasts between the sequences; note that with Koch's (1972) method, however, one first computes within-individual contrasts and then ranks them). There are several ways of ranking the observations in the context of survival analysis; Feingold & Gillespie (1996) employ Gehan's (1965a & b) scores; log-rank scores (see explanation in, e.g., Lawless, 1982, p. 420; Lee, 1992, p.109-112) might be preferable (Prentice & Marek, 1979; O'Brien & Fleming, 1987; Kalbfleisch & Prentice, 1980; Lee, 1992; Lawless, 1982). The p-value for this test could be obtained with a t-test, a Mann-Whitney test, or a randomization test. This method is easy to apply, and it can be used with multiple strata or trials composed of dual designs, e.g., by using the extended Mantel-Haenszel test with the log-ranked data (e.g., TK) or using randomization tests where randomization is constrained within strata. An example of the application of this method to a behavioural experiment is given in Díaz-Uriarte (1999). An alternative to Feingold & Gillespie's (1996) approach is to apply the methods in "Ordinal responses" to log-ranks of the data (see TK, p. 365).

5.7 Multivariate responses and repeated measures within periods

Behavioural ecology experiments frequently collect more than one response variable (e.g., in an anti-predator experiment in each period we might measure distance from the predator and time to re-emerge from the refuge, so we would have measured q=2 different response variables). This is somewhat similar to making repeated measurements (of the same response variable) within each time period (e.g., in each one of p periods, we might record the preferred perch height at 5 min intervals during 1 h; thus we have q "sub-periods" —here q=12— or different measurement occasions within each period). In both cases these are called "doubly multivariate" or "multivariate repeated measures". Multiple univariate tests of each one of the response variables (or at each one of the repeated observation times) can result in inferential problems as they ignore possible dependencies between observations (e.g., Krzanowski, 1990, p. 235 & ff.; Johnson & Wichern, 1998). Sometimes there is a large increase in Type I error rate (i.e., the true experiment-wise alpha level is larger than the nominal alpha level); other times fully multivariate approaches can attain larger power by using the information from the correlation among variables. With multiple responses it is frequently advised (e.g., Johnson & Wichern, 1998) that one should initially use a multivariate test and only if it reveals significant differences employ univariate tests on each response variable.

For metric data, JK devote a chapter (ch. 6) to repeated observations of the same variable. First, we could summarise the repeated data for each individual into one or a few statistics, such as area under the curve, slope and intercept, etc.; this is the simplest approach. However, this approach is problematic when the data are incomplete, and when covariates take different values during the observation session. Moreover, use of this approach requires obtaining a scientifically meaningful data summary, and thus assuming that all the information in the data that is not reflected by the summary statistic(s) is scientifically uninteresting (see also Crowder & Hand, 1990, ch. 1; Diggle et al., 1994, ch. 6 for discussion).

With two-sequences designs, a second approach (see JK, ch. 6) is to obtain individual contrasts (see above) for each sub-period q; thus, we reduce the data from a total of q*p to q derived measurements, and can analyse these q derived measurements with appropriate repeated measures techniques (e.g., MANOVA). For instance, Patel & Hearne (1980; see also Rodríguez-Carvajal & Freeman, 1999, p. 399) use a multivariate linear model and obtain, for each subject, a new transformed variable which is a linear combination of the original responses over the q sub-periods, and then use a two-sample Student's t-test on the transformed variable. This procedure tests the hypothesis that the sum of treatment effects over all periods is the same for the two sequences (and thus would not be appropriate with multiple responses —different variates).

A third approach, more satisfactory and flexible (and a necessity with more complicated designs) is to fit all the data in a single model (i.e., avoid reducing the data to q derived measurements). We can use a split-plot in time repeated measures ANOVA where we have three strata: between-subjects, within-subjects-among-periods, and within-period (i.e., the "sub-period" level). These analyses, like other split-plot-in-time repeated measures, make assumptions about the covariance structure which might not be appropriate; moreover, they are cumbersome if the spacing between successive measurements is unequal or if there are missing data. Thus we can also employ linear mixed effects models by specifying the corresponding random effects and covariance structures (see an example in Littell et al., 1996, pp. 388 and ff.). In addition, Galecki (1994) discusses some covariance structures which can be used with mixed models and allow flexibility for modelling the correlation structures for each repeated factor. These structures can be fitted using SAS's PROC MIXED; with the nlme library for S-Plus and R these structures can be fitted by defining the appropriate correlation structure.

With multiple response variables, application of Galecki's (1994) structures might not be appropriate (as they require that the marginal covariance structure associated with time be the same for every response variable). Thus, mixed models with more complex covariance

structures (and a larger number of parameters) need to be fitted (e.g., Amemiya, 1994; Vonesh & Chinchilli, 1997). These models could be fitted, for instance, using a completely unstructured (positive-definite) variance-covariance matrix (but in this case we would probably be estimating too many parameters). Alternatively, in S-Plus or R it might be possible to define special covariance structures tailored to our specific situation (e.g., unstructured except for blocks along the diagonal with particular structures for the within-variate covariance structure).

With categorical data, both GEE and generalized mixed models can accommodate multiple responses, although the latter requires that we specify the covariance structure. With time-to-event data, multiple responses can be easily analysed with the marginal approach of Lee et al (1992; we only need to obtain the quadratic form for the multivariate tests as in pp. 1066 and 1070 in Wei et al., 1989; see also documentation of library survival5) and the log-linear models of Lindsey et al. (1996; see pp. 531 and ff. for a worked example).

For some cross-over designs with multivariate normal responses, some simple approaches have been worked out. Rodríguez-Carvajal & Freeman (1999) show how to carry out a multivariate analysis in the 2x2 case using Hotelling's T^2 (a common statistic for multivariate comparisons of two groups; e.g., Morrison, 1990; Krzanowski, 1990). Grender & Johnson (1993; pp. 71-74 and 84) had proposed a similar but more general approach that can be extended to some higher-order designs, and it is applicable to both repeated measures and multiple responses, and to multiple responses with repeated measures for each response. The tests of Rodríguez-Carvajal & Freeman (1999) and Grender & Johnson (1993) for the multiple response situation is a simultaneous (multivariate) test of the hypothesis that the treatment effect vectors are the same in both sequences (which is appropriate when variates are not measured in the same scale), and differs from the test of Patel & Hearne (1980) explained above.

A different approach is to use nonparametric, rank-based, and randomization multivariate tests. Analogous to robust and nonparametric tests, the first step is to reduce the p*q measurements of each individual to a set of q variates by applying within-individual contrasts separately

to each variate (see Nonparametric section). We will refer to these as w-q. (With survival data a possibility is to apply the methods of Feingold & Gillespie (1996) by obtaining the w-q from the log-ranks or Gehan's scores of the data —not the original, censored, data; however, it is unknown how well this approach works). This first step of obtaining the w-q variates will be common to all the remaining multivariate tests. The next step is to compare, with the appropriate multivariate test, the w-q variates among sequences. Therefore, we can apply any multivariate test provided that we can set the hypothesis test as a comparison among sequences of within-individual contrasts. This will be possible (see JK, pp. 171 & ff.; SN, pp. 144-152; "Metric responses: nonparametric and robust methods" section) with two-treatment designs composed of pairs of dual sequences and with designs for more than two treatments that have the special structure in Table 4.5b, p. 101 in the previous chapter (but it might not be possible otherwise; this emphasises again the need to consider design and analysis before conducting the experiment). As was done before, we might want to start with within-individual contrasts that include carry-over effects, and later recompute the w-q from contrasts without carry-over if multivariate and univariate tests show no evidence of carry-over effects in any variable.

A very simple approach is to use the test in O'Brien (1984); first, each w-q is ranked separately; next, for each individual we compute Si as the sum of the ranks of all of the w-q. We test the null hypothesis of no overall difference between treatments by comparing the Si's between sequences, using a two-sample t-test, a rank-sum test, or a randomization test. This method can be extended to accommodate individual-level covariates (by using, e.g., a linear model with Si as the response and sequence and covariate as independent variables) and blocking (see "Metric responses: nonparametric and robust methods"). This application of O'Brien's test is very similar to Patel & Hearne's (1980) method, except that we use a linear combination of the ranks instead of the original variables (which is what makes it possible to apply the test to variables measured in different scales). A drawback of O'Brien's test is that it is appropriate only for some limited alternative hypotheses (see below).

After ranking each w-q separately, an alternative to O'Brien's (1984) test is to use the multivariate extension of the Kruskal-Wallis test (Puri & Sen, 1971, p. 184 & ff.), which is equivalent to applying a MANOVA on the separately ranked w-q variates (Zwick, 1985; note that with only two groups a MANOVA is the same as Hotelling's T^2). This is the test discussed in Johnson & Grender (1993; however they compute the test statistic using N*Pillai-Bartlett's trace, instead of (N-1)*Pillai-Bartlett's trace, as in Zwick, 1985; this is inconsequential if a randomization test is used, but not if the chi-square approximation is used).

A different test is obtained by applying the procedures of Mielke and collaborators (Mielke et al., 1976, 1981 a & b) to the w-q variates (without ranking), as explained in Johnson & Mercante (1996). This method does not assume any particular distribution for the data or homoscedasticity. We compute the average distance among the individuals of the two sequences in the q-dimensional space defined by the w-q variates, using an appropriate distance metric (e.g., Euclidean distance —but distance metric can affect power; Díaz-Uriarte & Nordheim, in prep.). Under the null hypothesis, permuting individuals randomly between the two sequences should have no effect on the average within-sequence distance, but under the alternative hypothesis permuting individuals should increase the average within-sequence distance. (P-values can be obtained from randomization tests, or using an approximation; see Mielke et al., 1976, 1981b; Berry & Mielke, 1983). When different response variables are measured in different scales, we will probably want to give equal weights to all variables; equal weights can be achieved by scaling the data (e.g., to a mean of zero and variance of one) before computing the within-subject comparisons or by applying the test to the ranks of the w-q variates —where each w-q is ranked separately—; (see Johnson & Mercante, 1996). An example of the application of this method to a behavioural study is given in Díaz-Uriarte (1999).

The tests discussed so far have been previously used with cross-over designs. Besides them, other randomization (e.g., Manly, 1997, ch.12; Edgington, 1995, ch. 8) and rank-based (e.g., Puri & Sen, 1971, 1985; Thompson, 1991; Choi & Marden, 1997; Hettmansperger et al., 1998)

multivariate tests could potentially be applied, either to the w-q variates or their ranks (with ranks computed for each variate separately or all together, depending on the test).

In summary, we can apply a fully multivariate approach to the original responses; this requires modelling the variance-covariance matrix in linear mixed models but not necessarily with GEE's or marginal survival models. When this is not feasible, multivariate and repeated measures tests can be applied to the w-q variates/responses. The latter, although more robust than, say, a fully multivariate linear mixed model, can also be considerably less powerful as we lose degrees of freedom when we reduce the data to w-q contrasts The appropriate statistic will depend on the null and alternative hypotheses and the structure of the data (and should not be decided based upon the results of the tests). For example, O'Brien's (1984) test is not designed to detect treatment effects that occur in only a few variates, or when the responses in different variates are not consistent (e.g., if there are negative correlations among variates). On the other hand, Hotelling's T^2 is not the most powerful test against restricted alternatives. Moreover, among nonparametric and rank-based multivariate tests, performance can be strongly affected by the shape of the distributions. Finally, different multivariate tests make different assumptions (normality, homoscedasticity, symmetry of distributions, etc.). Discussion can be found in Smith (1998), Choi & Marden (1997), Manly (1997, ch. 12), Edgington (1995, ch. 8), Westfall & Young (1993, ch. 6), Lachin (1992), Bernstein et al. (1988), and O'Brien (1984).

A different approach is to adjust the p-values to control for the increase in Type-I error rate from multiple univariate tests (e.g., Wright, 1992 and references therein; two articles in biological journals are Rice, 1989 and Chandler, 1995). These adjustments are better suited for situations (such as data snooping) where we are testing many individual hypotheses and want to control overall Type I error rates (e.g., we want to examine in which of five response variables a treatment has some effect), but are probably not the best approach when we conduct our experiment with the objective of testing a particular multivariate hypothesis (specified before the experiment was conducted); this approach is also useful when it is not possible to combine

the different tests into a single multivariate test. Most of the most recent methods (e.g., Hochberg's and Holm's sequential Bonferroni methods) provide much higher power than the traditional Bonferroni method (without increasing experiment-wise error rates), and some of them increase this power further by taking into account possible covariation among variables (e.g., Westfall & Young, 1993). For instance, the resampling-based methods in Westfall & Young (1993; see also SAS Institute, 1996, documentation for PROC MULTTEST) could be applied to the between sequence comparison of the w-q variates. Alternatively, we can employ the usual methods for cross-over trials with each variable independently, and later make an overall statement about the effect of a treatment by using, for example, Holm's multiple comparisons method.

Even in the absence of rigorous statistical methods for dealing with multiple response variables, some of the inferential problems arising from multiple responses can be minimised with careful experimental design and analysis. For instance, what hypotheses will be tested, and with what variables, can be specified a priori; also, different variables can be used to test different (biological) hypotheses, so that even if the data are not statistically independent, they at least refer to very different biological phenomena. This is not to suggest that other variables should not be examined for treatment effects, but just that testing of pre-specified hypotheses should be differentiated from hypotheses generation, for which data snooping might be well suited (see also discussion in Stewart-Oaten, 1995). Paraphrasing Rice (1989, p. 225), adjustment for multiple testing is necessary because, otherwise, as authors we will be spending many pages discussing spurious results, and as readers we will be wasting our time reading about relationships that can be explained just by chance.

5.7.1 PCA in lieu of MANOVA?

A potential mistake in the analysis of multiple responses is to try to use Principal Components Analysis (e.g., Morrison, 1990; Krzanowski, 1990; Bernstein et al., 1988) to reduce the

dimensionality of the response space, and then analyse the principal components scores as if they were independent response variables. This procedure is inappropriate for two reasons. First, if we want to reduce the dimensionality of the problem in the context of considering differences between groups, we should use canonical variates, which are different from principal components; canonical variates are closely related to MANOVA, canonical correlation, and discriminant analysis (see Krzanowski, 1990, p. 291-300 and 370-385; Bernstein et al., 1988, ch. 10; Digby & Kempton, 1987, pp. 75-77). Second, when using PCA we would be mixing within and among-individual covariation in the response variables. However, it should be possible to use canonical variate analysis on the w-q variates (including randomization-based canonical variate analysis —Manly, 1997, p. 274).

5.8 Plotting in cross-over designs

Plotting is a key tool in statistical analysis and can help us spot patterns and problems in the original data and the fitted models. We can plot the original data, plot some linear functions of the data, or make plots that are specific for the types of analyses carried out (particularly helpful to examine violations of model assumptions, such as residual plots). I will briefly review the first two here.

Initial plots of the data will help detect errors in the transcription or recording of data, and will give an idea of the results that could be expected. JK (p. 20) refer to **subject profile** plots where, for each sequence, the response of each subject is plotted over the different treatment periods, and the responses of each subject are connected with a line. These plots help identify period and treatment effects, potential outliers, and variation within and among sequences. For designs with more than two treatments, it is convenient to add treatment labels in the x-axis. In **treatment by treatment scatter-plots** (SN, p. 188), we plot each patient's values using each treatment response as a dimension.

The **response by patient scatterplot** (SN, p. 125 and 187) depicts the response variable (y-axis) by the sequence, using the same symbol across sequences to identify treatments; all the responses of a subject are shown in the same vertical line (x-axis position). This plot conveys a lot of information: variation within-subjects, variation among sequences, magnitude of differences between treatments, and possible differences in treatment effects across sequences (e.g., treatment*period interaction), as well as potential outliers (either a whole subject or observations within an otherwise non-outlying subject). This plot and the subject profiles plot complement each other, as they convey similar information in different ways. In these plots, covariates or other factors can be added by using symbols. Plots for time to event data are based on the survival function and are shown in Feingold & Gillespie (1996). Non-metric data are generally difficult to plot conveniently, and tables are probably more useful (but see SN, p. 188-190).

The second type of plots are those that depict some function of the data, such as the linear contrasts. These plots are very useful at the initial and intermediate stages of formal analyses. For the 2x2 design, JK (p. 28-30) discuss a plot that helps understand the **role of carry-over and treatment effects**. In a scatterplot, each individual's sum over the two periods is shown in the x-axis and each individual's difference between the first and second periods in the y-axis; individuals from each of the two sequences are plotted with different symbols, and the outermost points of each sequence are joined (i.e., we draw the convex hull of each sequence group). If there are only strong treatment effects, we will see two non-overlapping curves that are separated in the vertical direction; if there are carry-over effects, the separation will be along the horizontal axis. This plot also gives visual information on the variability in each sequence (for parametric analyses, variance should be the same in each group). The **groups-by-period plot** (JK, p. 20) shows the group by period means for each sequence, connected by a segment. These are very similar to the usual interaction plots in linear models. Plotting the linear contrast by a covariate can be particularly helpful to understand the role of continuous covariates. Miller (1999) has proposed two types of plots that help identify outliers

and indicate whether representing differences between samples by a single statistic (such as the mean) is appropriate; these plots allow us to examine subject by treatment interactions and changes in carry-over effect over time.

Summary plots of results should avoid two potential pitfalls. First, if analyses have been nonparametric it is misleading to use plots that represent a mean and its standard error, as these have no relationship with the actual analyses conducted (and could suggest that the mean and s.e. are adequate characterisations of the data distribution, which they are not). Second, in cross-over trials the estimator is based on within-individual differences, and the relevant source of variance is the within-individual variability, not the among-individual variability. Thus, a plot of the overall mean of treatments A and B, each with an standard error, would be of little use as the analyses were conducted using within individual differences; moreover, this plot can suggest no effect even when there is a strong one. Instead, it is preferable to plot the estimated treatment difference with its standard error (with no treatment differences, the confidence interval should cover 0). If we need to present the estimates of the actual responses with with some measure of variability, it is best if those treatment means are adjusted treatment means (as obtained from, e.g., linear models after correcting for effects of period and other fixed effects), and if a cautionary statement is added to the figure legend indicating that those means and s.e. cannot be used to conduct a visual test of the hypothesis.

5.9 Sample size and missing data

Discussion of sample size and power is provided in SN (211-219), Hills & Armitage (1979), and Ezzet & Whitehead (1992). Sample size calculations can be extremely complicated except for the simplest designs, and when planning trials we would need information on variances, which is not always available before the trial starts.

The consequences of missing data can be particularly serious for the 2x2 design; the simplest

strategy is to use only subjects without missing data, but other strategies are possible (JK, p. 76-80). For other designs, the consequences of missing data are not necessarily that serious, and probably all the available data from every subject should be used (see SN, p. 219-221; see also Low et al., 1999 for discussion of robustness of cross-over designs to dropouts).

It is important to understand what is the missing data mechanism (e.g., Diggle et al., 1994, ch. 11; Albert, 1999). A common classification is based on Littell & Rubin (1987). Data are missing completely at random (MCAR) if the missing mechanisms is independent of both the observed and actual missing value; they are missing at random (MAR) when the missing mechanism is independent of the actual missing value but depend on observed data (e.g., if it depends on previously observed values); and they are missing non-randomly (= informative missing mechanism or non-ignorable missingness) when the missing mechanisms depends on the values of the missed observations. For instance, suppose we are measuring fight duration in an experiment where each subject is scheduled to be observed five times per day, but occasionally we can not obtain complete records for each individual. If there is a constant probability that we cannot find the subject for the scheduled observation we have a MCAR mechanism. If, however, long previous fights make it more unlikely that we will able to find the subject for the following trial (e.g., following a long fight an animal is more likely to move somewhere else), then we have a MAR mechanism. We will have non-ignorable missingness if the probability that we observe a short fight is smaller than that of observing a long fight (i.e., the probability of recording a fight increases with fight duration, the variable we are measuring).

The statistical methods discussed above can accommodate MCAR data; some of them (e.g., linear mixed models, but not GEE) also accommodate MAR data; but most methods will be biased with informative missing values (e.g., experiments where the probability of having missing data depend on the treatment applied). Application of multivariate/repeated measures within periods techniques can be much more complicated in the presence of missing values or incomplete observations (see, e.g., Davis, 1991; Lachin, 1992; Palesch & Lachin, 1994).

5.10 Conclusions

Cross-over designs can be very useful in many behavioural experiments (see DU-1); however, their analyses are more complicated than those of parallel trials. When planning a cross-over trial we should consider both the design and analysis, as the type of response variable can affect the choice of design. Cross-over designs will be much easier to analyse if we can keep the design simple, minimising nesting and crossing of among-subject treatments (but if the setup does include these factors, they should be incorporated in the analyses).

Analysis of categorical data (specially ordered responses) can be complicated with cross-over designs, and generally requires at least moderate sample sizes ($\geq$10 individuals per sequence group); even with moderate sample size, power might be too low to detect small, but biologically relevant, differences between treatments. Analysis of time to event data can also be unsatisfactory, but is easier if censoring time is common for all individuals. More complex designs, such as those that include blocks and covariates, can make analysis of categorical and time to event data very complicated. Modifying the experimental protocol might ameliorate some of these problems; for example, to avoid censored data we might make observation periods longer, and to eliminate categories such as "low perch", "medium height", "high perch" we might be able to actually measure perch height. In particular, it is best to always obtain data at as high a level as possible in the measurement hierarchy (i.e., as close to interval as possible), and to remember that degrading data into categories such as orderings or 0/1 will make analyses more complicated. Experiments with three or more treatments are inherently more complicated to design and analyse, in particular if nonparametric and robust methods will be used. Experiments that measure multiple responses should use multivariate techniques. Finally, how carry-over and period effects are dealt with should be made explicit.

5.11 Acknowledgements

C. Lázaro-Perea for advice and comments on the ms. A. R. Ives, B. Jones, J. K. Lindsey, C. A. Marler, and C. T. Snowdon for comments on the ms. E. V. Nordheim for statistical discussion. A. T. Galecki for discussion on multivariate mixed-effects models. Manuscript preparation partially supported by a John and Virginia Emlen Graduate Student Fellowship, Dept. Zoology, University of Wisconsin-Madison.

5.12 References

Agresti, A. 1990. Categorical Data Analysis. New York: John Wiley & Sons.

Agresti, A. 1993. Distribution-free fitting of logit models with random effects for repeated categorical responses. Statistics in Medicine, 12, 1969-1987.

Albert, P. S. 1999. Longitudinal data analysis (repeated measures) in clinical trials. Statistics in Medicine, 18, 1707-1732.

Allison, P. D. 1995. Survival Analysis Using the SAS System: a Practical Guide. Cary, NC: SAS Institute Inc.

Amemiya, Y. 1994. On multivariate mixed model analysis. In: Multivariate analysis and its applications. IMS lecture notes, vol. 24. (Ed. by T. W. Anderson, K. T. Fang, and I. Olkin), pp. 83-95. Hayward, CA: Institute of Mathematical Statistics

Aragaki, D. L. R. and Meffert, L. M. 1998. A test of how well the repeatability of courtship predicts its heritability. Anim. Behav., 55, 1141-1150.

Becker, M. P. and Balagtas, C. C. 1993. Marginal modeling of binary cross-over data. Biometrics, 49, 997-1009.

Bellavance, F. and Tardif, S. 1995. A nonparametric approach to the analysis of three-treatment three-period crossover designs. Biometrika, 82, 865-875.

Bennington, C. C. and Thayne, W. V. 1994. Use and misuse of mixed model analysis of variance in ecological studies. Ecology, 75, 717-722.

Bernstein, I. H., Garbin, C. P., and Teng, G. K. 1988. Applied Multivariate Analysis. New York: Springer-Verlag.

Berry, K. J. and Mielke, P. W. Jr. 1983. Computation of finite population parameters and approximate probability values for multi-response permutation procedures (MRPP). Commun. Statist.-Simula. Computa., 12, 83-107.

Brown, H. K. and Kempton, R. A. 1994. The application of REML in clinical trials. Statistics in Medicine, 13, 1601-1617.

Brunner, E. and Neumann, N. 1987. Non-parametric methods for the 2-period cross-over design under weak model assumptions. Biometrical Journal, 29, 907-920.

Chandler, C. R. 1995. Practical considerations in the use of simulatenous inference for multiple tests. Animal Behaviour, 49, 524-527.

Choi, K. and Marden, J. 1997. An approach to multivariate rank tests in multivariate analysis of variance. Journal of the American Statistical Association, 92, 1581-1590.

Collett, D. 1994. Modelling Survival Data in Medical Research. London: Chapman & Hall.

Conover, W. J. 1980. Practical Nonparametric Statistics. New York: John Wiley & Sons.

Crawley, M. J. 1993. GLIM for Ecologists. Oxford: Blackwell.

Crowder, M. J. and Hand, D. J. 1990. Analysis of Repeated Measures. New York: Chapman & Hall.

Crowley, P. H. 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. Ann. Rev. Ecol. Syst., 23, 405-447.

Davidian, M. and Giltinan, D. M. 1995. Nonlinear Models for Repeated Measurement Data. London: Chapman & Hall.

Davis, C. S. 1991. Semi-parametric and non-parametric methods for the analysis of repeated measurements with applications to clinical trials. Statistics in Medicine, 10, 1959-1980.

DeWitt, T. J., Sih, A., and Hucko, J. A. 1999. Trait compensation and cospecialization in a freshwater snail: size, shape and antipredator behaviour. Anim. Behav., 58, 397-407.

Díaz-Uriarte, R. 1999. Anti-predator behaviour changes following an aggressive encounter in the lizard *Tropidurus hispidus*. Proceedings Royal Society of London, Series B, 266, 2457-2464.

Digby, P. G. N. and Kempton, R. A. 1987. Multivariate Analysis of Ecological Communities. London: Chapman and Hall.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. 1994. Analysis of Longitudinal Data. Oxford: Oxford University Press.

Dobson, A. J. 1990. An Introduction to Generalized Linear Models. London : Chapman & Hall.

Donev, A. N. 1998. Crossover designs with correlated observations. Journal of Biopharmaceutical Statistics, 8, 249-62.

Ducrocq, V. 1997. Survival analysis, a statistical tool for longevity data. 48th Annual Meeting European Association for Animal Production, 25-28 August, 1997, Vienna, Austria.

Edgington, E. S. 1995. Randomization Tests, 3rd Ed. New York: Marcel Dekker.

Elswick, R. K. and Uthoff, V. A. 1989. A nonparametric approach to the analysis of the 2-treatment, 2-period, 4-sequence crossover model. Biometrics, 45, 663-667.

Ezzet, F. and Whitehead, J. 1991. A random effects model for ordinal responses from a crossover trial. Statistics in Medicine, 10, 901-6.

Ezzet, F. and Whitehead, J. 1992. A random effects model for binary data from crossover clinical-trials. Applied Statistics-Journal of the Royal Statistical Society Series C, 41, 117-126.

Ezzet, F. and Whitehead, J. 1993. A random effects model for ordinal responses from a crossover trial - Reply. Statistics in Medicine, 12, 2150-2151.

Feingold, M. and Gillespie, B. W. 1996. Cross-over trials with censored data. Statistics in Medicine, 15, 953-967.

Fidler, V. 1984. Change-over clinical trial with binary data: mixed-model-based comparison of tests. Biometrics, 40, 1063-1070.

France, L. A., Lewis, J. A., and Kay, R. 1991. The analysis of failure time data in crossover studies. Statistics in Medicine, 10, 1099-1113.

Galecki, A. T. 1994. General class of covariance structures for two or more repeated factors in longitudinal data analysis. Commun. Statist.-Theory Meth., 23, 3105-3119.

Gehan, E. A. 1965. A generalized two-sample Wilcoxon test for doubly censored data. Biometrika, 52, 650-653.

Gehan, E. A. 1965. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika, 52, 203-223.

Good, P. 1994. Permutation Tests. New York: Springer-Verlag.

Grender, J. M. and Johnson, W. D. 1993. Analysis of crossover designs with multivariate response. Stats. Medicine, 12, 69-89.

Hafner, K. B., Koch, G. G., and Canada, A. T. 1988. Some analysis strategies for three-period changeover designs with two treatments. Statistics in Medicine, 7, 471-481.

Heimann, G. and Neuhaus, G. 1998. Permutational distribution of the log-rank statistic under random censorship with applications to carcinogenicity assays. Biometrics, 54, 168-184.

Hettmansperger, T. P., Mottonen, J., and Oja, H. 1998. Affine invariant multivariate rank tests for several samples. Statistica Sinica, 8, 785-800.

Hills, M. and Armitage, P. 1979. The two-period cross-over clinical trial. British Journal of Clinical Pharmacology, 8, 7-20.

Horton, N. J. and Lipsitz, S. R. 1999. Review of software to fit Generalized Estimating Equation regression models. The American Statistician, 53, 160-169.

Johnson, D. H. 1995. Statistical sirens: the allure of nonparametrics. Ecology, 76: 1998-2000.

Johnson, W. D. and Grender, J. M. 1993. Multivariate nonparametric analysis for the two-period crossover design with application in clinical trials. Journal of Biopharmaceutical Statistics, 3, 1-12.

Johnson, W. D. and Mercante, D. E. 1996. Analyzing multivariate data in crossover designs using permutation tests. Journal of Biopharmaceutical Statistics, 6, 327-42.

Johnson, R. A. and Wichern, D. W. 1998. Applied Multivariate Statistical Analysis, 4th ed. N.J., Prentice Hall.

Jones, B. and Kenward, M. G. 1989. Design and Analysis of Cross-Over Trials. New York: Chapman & Hall.

Jung, J. W. and Koch, G. G. 1999. Multivariate non-parametric methods for Mann-Whitney statistics to analyse cross-over studies with two treatment sequences. Statistics in Medicine, 18, 989-1017.

Kalbfleisch, J. D. and Prentice, R. L. 1980. The Statistical Analysis of Failure Time Data. New York: John Wiley & Sons.

Kenward, M. G. and Jones, B. 1991. The analysis of categorical-data from cross-over trials using a latent variable model. Statistics in Medicine, 10, 1607-1619.

Kenward, M. G. and Jones, B. 1994. The analysis of binary and categorical data from crossover trials. Statistical Methods in Medical Research, 3, 325-44.

Klein, J. P. and Moeschberger, M. L. 1997. Survival Analysis. New York: Springer-Verlag.

Koch, G. G. 1972. The use of non-parametric methods in the statistical analysis of the two-period change-over design. Biometrics, 28, 577-584.

Koch, G. G. and Edwards, S. 1988. Clinical efficiency trials with categorical data. In: Biopharmaceutical statistics for drug development. (Ed. by K. E. Peace), pp. 403-457. New York: Marcel Dekker

Koch, G. G., Gitomer, S. L., and Skalland, L. 1983. Some non-parametric and categorical data analyses for a change-over design study and discussion of apparent carry-over effects. Statistics in Medicine, 2, 397-412.

Krzanowski, W. J. 1990. Principles of Multivariate Analysis. New York: Oxford University Press.

Lachin, J. M. 1992. Some large-sample distribution-free estimators and tests for multivariate partially incomplete populations. Statistics in Medicine, 11, 1151-1170.

Lawless, J. F. 1982. Statistical Models and Methods for Lifetime Data. New York: John Wiley & Sons.

Lee, E. T. 1992. Statistical Methods for Survival Data Analysis. New York: John Wiley & Sons, Inc.

Lee, E. W., Wei, L. J., and Amato, D. A. 1992. Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In: Survival analysis: state of the art. (Ed. by J. P. Klein and P. K. Goel), pp. 237-247. The Netherlands: Kluwer Academic

Liang, K.-Y., Zeger, S. L., and Qaqish, B. 1992. Multivariate regression analyses for categorical data. J. R. Statist. Soc. B , 54, 3-40.

Lin, D. Y. 1993. MULCOX2: a general computer program for the Cox regression analysis of multivariate failure time data. Comp. Meth. Progr. Biom., 40, 279-293.

Lin, D. Y. 1994. Cox regression analysis of multivariate failure time data: the marginal approach. Statistics in Medicine, 13, 2233-2247.

Lindsey, J. K. 1993. Models for Repeated Measurements. Oxford: Clarendon Press.

Lindsey, J. K., Jones, B., and Lewis, J. A. 1996. Analysis of cross-over trials for duration data. Statistics in Medicine, 15, 527-35.

Lipsitz, S. R., Kim., K., and Zhao, L. 1994. Analysis of repeated categorical data using generalized estimating equations. Statistics in Medicine, 13, 1149-1163.

Littell, R. C., Freund, R. J., and Spector, P. C. 1991. SAS System for Linear Models, 3rd Ed. Cary, NC: SAS Institute.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. 1996. SAS System for Mixed Models. Cary, NC: SAS Institute.

Littell, R. J. A. and Rubin, D. B. 1987. Statistical Analysis With Missing Data. New York: John Wiley & Sons.

Low, J. L., Lewis, S. M., and Prescott, P. 1999. Assessing robustness of crossover designs to subjects dropping out. Statistics and Computing, 9, 219-227.

Ludbrook, J. and Dudley, H. 1998. Why permutation tests are superior to t and F tests in biomedical research. The American Statistician, 52, 127-132.

Manly, B. F. J. 1997. Randomization, Bootstrap, and Monte Carlo Methods in Biology, 2nd Ed. London: Chapman & Hall.

Maritz, J. S. 1995. Distribution-Free Statistical Methods, 2nd Ed. London: Chapman & Hall.

McCullagh, P. and Nelder, J. A. 1989. Generalized Linear Models, 2nd Ed. New York: Chapman & Hall.

Mielke, P. W. Jr., Berry, K. J., and Johnson, E. S. 1976. Multi-response permutation procedures for a priori classifications. Commun. Statist.-Theor. Meth., A5, 1409-1424.

Mielke, P. W. Jr., Berry, K. J., and Brier, G. W. 1981a. Application of multi-response permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. Monthly Weather Review, 109, 120-126.

Mielke, P. W., Berry, K. J., Brockwell, P. J., and Williams, J. S. 1981b. A class of nonparametric tests based on multiresponse permutation procedures. Biometrika, 68, 720-724.

Miller, W. E. 1999. A visualization of cross-over data using linear functions. Statistics in Medicine, 18, 975-987.

Morrison, D. F. 1990. Multivariate Statistical Methods, 3r Ed. New York: McGraw-Hill.

Noreen, E. W. 1989. Computer-Intensive Methods for Testing Hypotheses: an Introduction. New York: John Wiley & Sons.

O'Brien, P. C. 1984. Procedures for comparing samples with multiple endpoints. Biometrics, 40, 1079-1087.

O'Brien, P. C. and Fleming, T. R. 1987. A paired Prentice-Wilcoxon test for censored paired data. Biometrics, 43, 169-180.

Ohrvik, J. 1998. Nonparametric methods in crossover trials. Biometrical Journal, 40, 771-789.

Palesch, Y. Y. and Lachin, J. M. 1994. Asymptotically distribution-free multivariate rank-tests for multiple samples with partially incomplete observations. Statistica Sinica, 4, 373-387.

Patel, H. I. and Hearne III, E. M. 1980. Multivariate analysis for the two-period repeated measures crossover design with application to clinical trials. Coomu. Statist.-Theor. Meth., A9, 1919-1929.

Peace, K. E. and Koch, G. G. 1993. Statistical methods for a three-period crossover design in which high dose cannot be used first. J. Biopharmaceutical Statistics, 3, 103-116.

Pinheiro, J. C & Bates, D. M. 2000. Mixed-Effects Models in S. Springer-Verlag, in press.

Prentice, R. L. and Marek, P. 1979. A qualitative discrepancy between censored data rank tests. Biometrics, 35, 861-867.

Puri, M. L. and Sen, P. K. 1971. Nonparametric Methods in Multivariate Analysis. New York: John Wiley & Sons.

Puri, M. L. and Sen, P. K. 1985. Nonparametric Methods in General Linear Models. New York: John Wiley & Sons.

Ratkowsky, D. A., Evans, M. A., and Alldredge, J. R. 1993. Cross-Over Experiments: Design, Analysis, and Application. New York: Marcel Dekker.

Rice, W. R. 1989. Analyzing tables of statistical tests. Evolution, 43, 223-225.

Rodriguez-Carvajal, L. A. and Freeman, G. H. 1999. Multivariate AB-BA crossover design. J. Applied Statistics, 26, 393-403.

SAS Institute Inc. 1996. SAS/STAT Software: Changes and Enhancements Through Release 6.11. Cary, NC: SAS Institute.

Seaman, J. W. and Jaeger, R. G. 1990. Statisticae dogmaticae: a critical essay on statistical practice in ecology. Herpetologica, 46, 337-346.

Segal, M. R. and Neuhaus, J. M. 1993. Robust inference for multivariate survival-data. Statistics in Medicine, 12, 1019-1031.

Senn, S. 1993a. Cross-Over Trials in Clinical Research. New York: John Wiley & Sons.

Senn, S. 1993b. A random effects model for ordinal responses from a crossover trial [letter; comment]. Statistics in Medicine, 12, 2147-51.

Shen, C. D. and Quade, D. 1983. A randomization test for a three-period three-treatment crossover experiment. Commun. Statist.-Simula. Computa., 12, 183-199.

Smith, E. P. 1998. Randomization methods and the analysis of multivariate ecological data. Environmetrics, 9, 37-51.

Stewart-Oaten, A. 1995. Rules and judgments in statistics: three examples. Ecology, 76, 2001-2009.

Therneau, T. and Grambsch, P. 2000. Modeling Survival Data: Extending the Cox Model. Springer-Verlag, New York, in press.

Thompson, G. L. 1991. A unified approach to rank tests for multivariate and repeated measures designs. Journal of the American Statistical Association, 86, 410-419.

Tudor, G. and Koch, G. G. 1994. Review of nonparametric methods for the analysis of crossover studies. Statistical Methods in Medical Research, 3, 345-81.

Verbeke, G. and Molenberghs, G (eds).1997. Linear mixed models in practice. A SAS-Oriented approach. New York: Springer-Verlag

Vonesh, E. F. and Chinchilli, V. M. 1997. Linear and Nonlinear Models for the Analysis of Repeated Measurements. New York: Marcel Dekker.

Wei, L. J., Lin, D. Y., and Weissfeld, L. 1989. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association, 84, 1065-1073.

Westfall, P. H. and Young, S. S. 1993. Resampling-Based Multiple Testing. New York: John Wiley & Sons.

Woolson, R. F. and O'Gorman, T. W. 1992. A comparison of several tests for censored paired data. Statistics in Medicine, 11, 193-208.

Wrigth, S. P. 1992. Adjusted P-values for simultaneous inference. Biometrics, 48, 1005-1013.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. 1988. Models for longitudinal data: a generalized estimating equation approach. Biometrics, 44, 1049-1060.

Zwick, R. 1985. Nonparametric one-way multivariate analysis of variance: a computational approach based on the Pillai-Bartlett trace. Psychological Bulletin, 97, 148-52.