Evolutionary Ecology of Social Interactions Among Plants

Jay Michael Biernaskie

Doctor of Philosophy

Ecology and Evolutionary Biology
University of Toronto

2010

Abstract

Neighbouring plants can interact strongly, competing for resources including light, water, animal mutualists, and local germination sites. From an evolutionary perspective, this implies that a plant’s best resource acquisition strategy will usually depend on the traits of its neighbours, and for plants in particular, neighbours are often genealogical relatives. Here, I use a combination of theory and experiments to expose some important consequences of social interactions among plants. The first model analyzes selection on traits used to attract pollinators, showing that competitive interactions (in the absence of local relatedness) can select for exaggerated secondary sexual characters. To complement this model, I performed experiments that confirm the mechanisms by which adaptive pollinator foraging naturally leads to interactions among plants. The observed foraging behaviour (of bumble bees) also provides unique evidence for ‘Bayesian foraging’, a sophisticated type of resource assessment that depends on prior experience in a particular environment. A second model considers how selection on the sex allocation of cosexual, animal-dispersed plants leads to competition and cooperation over local germination sites, sometimes leading to the origin of gender dimorphism. The model reveals novel ecological contexts in which disruptive selection on sex allocation can arise, and in general, illustrates how selection for cooperation can facilitate or inhibit evolutionary diversification. In the models considered here, cooperation is indiscriminant, but plants might
also assess the relatedness of neighbours and cooperate with kin over non-kin. In the final chapter, I present experimental evidence that is consistent with preferential cooperation over soil resources among sibling plants. This study is the first to link a potentially cooperative resource allocation strategy with an increase in the mean fitness of related plants.
Acknowledgments

To begin with my longest list of thanks, I acknowledge the friends and colleagues who have commented on my manuscripts or otherwise facilitated my research. For reviewing manuscripts, I thank Rob Gegear, Patsy Haccou, Randy Mitchell, Michael Otterstatter, Jennifer Perry, Claus Rueffler, Mario Vallejo-Marín, Steve Walker, and a number of anonymous reviewers. The experiments in Chapters 3 and 4 were carried out in the laboratory of James Thomson, and I thank him for generously offering his space and resources. For advice on various issues that arose in Chapter 4, I thank Richard Green, Jan van Gils, and Eric Wajnberg. The experiment in Chapter 6 was performed in the Ramsay Wright greenhouse, where I was helped greatly by Bruce Hall and Andrew Petrie. Jim Dix and Trung Luu also helped enormously by building a large network of stakes for the greenhouse. John Stinchcombe, Anna Simonsen, and Brandon Campitelli provided valuable advice on the design of said experiment, and J. Stinchcombe very kindly donated the seeds; without this help, the experiment would not have been possible.

I want to give special thanks to my thesis committee, Peter Abrams, Aneil Agrawal, and Locke Rowe, particularly for the freedom and encouragement to follow my research interests. I hope that my work and our meetings were an interesting distraction. To P. Abrams in particular, I thank you for always being quick to read my work, for generously supporting my travels and expenses, and, more generally, for being an excellent mentor.

I was initially funded by a Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC). Later, I was funded by NSERC grants to P. Abrams and by scholarships and fellowships from the EEB department. I thank all of these sources for their support.

Finally, I thank my family for encouraging and supporting my career. My parents funded the majority of my early education, and this has certainly made it easier for me to complete my studies and to live comfortably while doing so. Life has also been easy because of the support and companionship of my best friend, Jennifer, who I am now proud to call my wife.
Statement of contributions

I was the lead author on all of the chapters in this thesis, but some of the chapters were collaborative projects that resulted in multi-authored publications. Some of the ideas in Chapter 2 are the product of discussions with Elizabeth Elle, my past advisor. I derived the models and wrote the paper myself, however, while in Toronto. I conceived the experiments in Chapter 3, but they were performed with the help of Rob Gegear, whose expertise was critical to the success of the project. Together, Rob and I decided on the design of the follow-up experiment in Chapter 4, and again, we collected most of the data. I also collected some of the data with a third collaborator, Steve Walker, and Steve and I were most involved in the writing of the paper. Together, Steve and I developed the optimal foraging and statistical models, but Steve programmed the models and produced the figures. With input from both Steve and Rob, I wrote the majority of the paper, although Steve wrote the Appendix and some sections of the methods. In the published version of Chapter 4, S. Walker and I declare equal contributions. Chapters 5 and 6 were conceived of and written by myself.
## Table of Contents

Acknowledgments ........................................................................................................................................ iv

Table of Contents ...................................................................................................................................... vi

List of Tables ........................................................................................................................................... ix

List of Figures ........................................................................................................................................ x

Chapter 1 General Introduction .................................................................................................................. 1

1 General Introduction ............................................................................................................................... 1

1.1 The Social Traits of Plants ................................................................................................................ 1

1.2 The Tragedy of the Commons ............................................................................................................. 2

1.3 Sex Allocation ................................................................................................................................. 4

1.4 Evolutionary Ecology of Social Interactions Among Plants ......................................................... 4

1.5 Literature Cited .............................................................................................................................. 6

Chapter 2 Competition and facilitation in pollinator attraction, and a theory for exaggerated secondary sexual traits in plants .................................................................................................................. 8

2 Abstract .................................................................................................................................................. 8

2.1 Introduction ...................................................................................................................................... 8

2.2 Models .............................................................................................................................................. 12

2.3 Discussion ....................................................................................................................................... 17

2.4 Literature Cited ............................................................................................................................... 21

2.5 Appendix: Investment in pollinator attraction in a simple patch-structured population ............. 26

2.6 Figures ............................................................................................................................................. 28

Chapter 3 Habitat assessment ability of bumble bees implies frequency dependent selection on floral rewards and display size ........................................................................................................ 31

3 Abstract ............................................................................................................................................... 31

3.1 Introduction .................................................................................................................................... 31

3.2 Materials and Methods .................................................................................................................. 34
List of Tables

Table 3.1 .......................................................................................................................................... 50
Table 6.1 ........................................................................................................................................ 137
Table 6.2 ........................................................................................................................................ 138
List of Figures

Figure 2.1 ..................................................................................................................... 28
Figure 2.2 ..................................................................................................................... 29
Figure 2.A.1 ................................................................................................................ 30
Figure 3.1 ..................................................................................................................... 47
Figure 3.2 ..................................................................................................................... 48
Figure 4.1 ..................................................................................................................... 75
Figure 4.2 ..................................................................................................................... 77
Figure 4.3 ..................................................................................................................... 79
Figure 4.A.1 ............................................................................................................... 80
Figure 4.A.2 ............................................................................................................... 81
Figure 5.1 .................................................................................................................... 108
Figure 5.2 .................................................................................................................... 110
Figure 5.3 .................................................................................................................... 111
Figure 5.4 .................................................................................................................... 113
Figure 5.5 .................................................................................................................... 114
Figure 5.A.2 ............................................................................................................. 116
Figure 6.A.1 ............................................................................................................. 131
Figure 6.1 .................................................................................................................... 132
Figure 6.2 .................................................................................................................... 133
Figure 6.3 .................................................................................................................... 134
Figure 6.4 .................................................................................................................... 135
Figure 6.5 .................................................................................................................... 136
“The lack of any plant equivalent to an animal society severely limits how far ‘sociobotany’ can go.”

J. Silvertown (1989); New Phytologist 112: 170

1 General Introduction

1.1 The Social Traits of Plants

Sociality can be defined simply as group living (Frank 2007). A universal consequence of living in groups is that, on the one hand, competition over local resources can be exacerbated, while on the other, group members can cooperate to enhance the productivity of the group as a whole (relative to other groups). This fundamental social tension applies to groups of plants just as it does to groups of animals (e.g., Wilson 1975), microorganisms (West et al. 2006), and all other levels of biological organization (Maynard Smith and Szathmary 1995). The following thesis is a contribution to the field of sociobotany; that is, it studies the ecology and evolution of competitive and cooperative interactions within plant groups.

Despite the immobility of their adult stage, plants can interact in a number of ways. The most obvious social interactions involve the exploitation of local resources—traits that determine a plant’s ability to compete for resources clearly influence the fitness of the individual itself and the fitness of its neighbours. Commonly studied traits of this kind include: plant height, used to acquire light (e.g., Falster and Westoby 2003; Weiner 1990); root proliferation, used to acquire soil resources (e.g., Gersani et al. 2001; Zhang et al. 1999); and the ‘aggressiveness’ of water use (Zea-Cabrera et al. 2006). It may be less obvious that plants also interact through the behaviour of animal mutualists. The behaviour of pollinators towards a plant with a particular quality of rewards, for example, may often depend on the reward quality of neighbouring plants (e.g., Klinkhamer et al. 2001). It follows that if there are fitness effects associated with the
pollinators’ behaviour, then the best trait for a plant to adopt will depend on the traits of its neighbours. This scenario and others like it can be conceptualized as an evolutionary game (see Vincent and Brown 2005), and indeed, the references cited above all fall within this framework. It is less popular to describe the same contexts (and the traits involved) as ‘social’, yet as I will show below, the analysis of natural selection acting on social traits is often a simple extension of standard game-theoretic approaches.

Social traits are categorized by their fitness effects on a focal individual bearing the trait and their fitness effects on neighbours (Hamilton 1964; Trivers 1985; West et al. 2007). Selfish traits, which have a positive impact on the focal individual but a negative effect on neighbours, are easiest to explain because natural selection tends to increase the relative fitness of individuals. Most of the plant traits cited above have been studied in a competitive context, where selection is predicted to favour strategies that further the selfish interests of individuals. Adaptive cooperation, where selection favours a trait that positively affects an individual’s neighbours, requires a comparatively special explanation. Here, I explain the evolution of cooperation as a solution to the ‘tragedy of the commons’ (see Hardin 1993), which is a common theme that arises throughout the thesis.

1.2 The Tragedy of the Commons

The tragedy of the commons analogy applies to most social interactions among plants and highlights the fundamental social tension between individual selfishness and group productivity (Frank 1998). As with most game-theoretic treatments of plant strategies, an analysis of selection in this context begins with an expression for the ‘neighbour-modulated’ fitness (Hamilton 1964) of an individual plant that expresses a variant trait value \(x\). The fitness, \(w\), of a focal individual with selfish tendency \(x\) in a group with average selfish tendency \(y\) can be written as

\[
w(x, y) = \frac{x}{y} (1 - y)
\]
(Frank 1995), where $x/y$ is the relative success of the focal individual, and $(1-y)$ is the productivity of the group (higher average selfishness results in lower productivity). A key feature of the models used throughout this thesis is that the focal individual’s trait value may be correlated with the group’s average trait value (and hence $y$ is a function of $x$). Given this relationship, the marginal change in the focal individual’s fitness with respect to $x$ is

$$\frac{dw}{dx} = \frac{\partial w}{\partial x} + \frac{\partial w}{\partial y} \frac{dy}{dx},$$

where, in the second step, the partial derivatives are evaluated at the equilibrium trait value ($y^*$) and $dy/dx$ is interpreted as a type of relatedness coefficient (Frank 1998). The solution to the equilibrium is $y^* = 1 - r$, showing that cooperation is favoured only when there is a positive correlation among the traits of neighbours. All individuals benefit from being in a cooperative group, but when $r$ is low, selfish competition escalates to the detriment of all. When $r$ is high, however, an individual can do best by cooperating because its neighbours will tend to show similar restraint. Notice that this neighbour-modulated approach (developed by Frank 1998; Taylor and Frank 1996) differs in perspective from Hamilton’s (1964) inclusive fitness approach (though both are mathematically equivalent). The neighbour-modulated approach considers social effects flowing from social partners to the focal individual, whereas the inclusive fitness approach considers effects flowing from a focal individual to genetically related social partners.

Indeed, the most important reason for neighbouring plants to have positively correlated trait values is due to genealogical relatedness. Many authors have noted that neighbouring plants are especially likely to be related because of limited dispersal, clonal growth and self-pollination (e.g., de Jong and Klinkhamer 2005; Wade 1980). When this is true, natural selection should sometimes favour cooperative plant traits such as short stature, restrained root proliferation, and prudent water use. Notwithstanding some important exceptions, however, analyses of the evolutionary consequences of social interactions among relatives have contributed little to current understanding of plant adaptations.
1.3 Sex Allocation

One of the important areas in which social interactions among plants have not been neglected is in the study of sex allocation (the allocation of reproductive resources to male and female functions; Charnov 1982; West 2009). Sex allocation is a social trait because the reproductive success of an individual through male or female function naturally depends on the sex allocation of interacting individuals. In fact, sex allocation is often recognized as the most important topic in the development of social evolution theory as a whole (e.g., Frank 1998; West 2009). Hamilton’s (1967) famous theory for local mate competition, for example, shows that when males (pollen) compete for mates locally but females (fruits) disperse from a local group, a female-biased strategy (restraint from local competition) is only favoured when social partners show similar restraint (i.e., relatedness is high), thus maximizing group productivity. Such insights from sex allocation theory have been successfully applied to understand many interesting plant reproductive strategies, including the major transitions in sexual systems (reviewed by deJong and Klinkhamer 2005). However, there still remain a number of open problems that might be solved by a more careful consideration of the full social context of plant sex allocation.

1.4 Evolutionary Ecology of Social Interactions Among Plants

This thesis is a collection of studies related to competitive and cooperative interactions among plants, each one touching on the ideas introduced above. The chapters, summarized here, appear in the following order:

Chapter 2. I suggest that competition among plants for pollinator visits can represent an important indirect interaction that is absent from most models of investment in pollinator rewards. Including this competition in simple models of sex allocation can escalate investment in rewards at the expense of allocation to gametes or at the expense of individual viability. The purpose of the model is to show that selection for exaggerated secondary sexual traits in plants must involve a type of social interaction. I also show that local relatedness can have the effect of favouring restraint in floral reward production as a form of cooperation.
Chapter 3. If pollinator residence (number of flowers probed) on a particular plant depends on the average reward quality of other plants in the population, results from Chapter 2 predict that selection can act to further escalate investment in rewarding traits. Experiments with bumble bees show that their residence on a plant with a particular reward value does indeed depend on learning the average reward quality of the habitat. This finding illuminates an important and largely neglected interaction among plants that has interesting implications for the evolution of plant reproductive strategies.

Chapter 4. Building on the previous chapter, experiments are used to test bumble bees’ ability to learn the distribution of rewards in a population. We use optimal foraging models and some unique methods for comparing theory with data to demonstrate an impressive match between bees’ behaviour and that of an optimal ‘Bayesian forager’. The chapter’s main focus is foraging ecology; implications for selection on plant traits are also briefly discussed.

Chapter 5. I shift again to sex allocation theory for cosexual plants. Here, I explore a unique problem in which allocation to female function (fruits) is coupled to the female dispersal rate via the preferences of fruit dispersers. A central goal is to understand the ecological and social contexts that can lead to disruptive selection on sex allocation. Analogies to the tragedy of the commons arise, where germination sites act as the local resource. Interestingly, I show that selection for relaxed competition over local resources can sometimes prevent disruptive selection and sometimes facilitate it.

Chapter 6. Finally, I report an experiment that asks whether groups of non-sibling plants suffer a ‘tragedy’ of competition that is relatively relaxed in sibling groups. In this case, any evidence for cooperation among sibs would imply plants’ ability to assess the relatedness of neighbours and respond accordingly. Overall, groups of siblings and non-siblings did not differ in mean productivity. Among the most productive groups, however, sibling groups tended to have higher mean fitness but smaller roots than non-sibs. This result suggests the possibility that sibling plants restrict root growth to minimize the tragedy of the commons.
1.5 Literature Cited


Chapter 2
Competition and facilitation in pollinator attraction, and a theory for exaggerated secondary sexual traits in plants

2      Abstract
Models are presented to explore the adaptive investment in pollinator rewards when the attractiveness of a particular trait depends on the mean trait value of conspecifics. In the main text, we focus on competitive interactions among non-relatives, where a plant’s attractiveness declines with the average trait value of competitors. Including this asymmetric competition in a standard model of hermaphroditic sex allocation shifts additional allocation to pollinator rewards at the expense of allocation to pollen and seeds. In a second model, plants can invest additional resources in pollinator rewards but suffer reduced viability and rising costs due to excess pollen removal and within-plant selfing (geitonogamy). Despite these accumulating costs, increasing the magnitude of asymmetric competition exaggerates the ESS investment in rewards beyond the case in which attractiveness depends only on a plant’s absolute reward value. We suggest that the type of frequency-dependent selection considered here is fundamentally equivalent to sexual selection in animal populations. In the appendix, it is shown that interactions among related plants can mitigate selection for exaggerated secondary sexual traits.

2.1    Introduction
The scope of sexual selection is often extended beyond Darwin’s (1871) explanation for exaggerated secondary sexual characters (see Clutton-Brock, 2004). For some, the theory encompasses any trait with a positive influence on mating success (e.g. Murphy, 1998)—a viewpoint that is particularly evident in the application of sexual selection theory to flowering

plants. Sexual selection in plants is typically defined as selection arising from differences in mating success (e.g. Arnold, 1994; Willson, 1994) or as selection for mating success that acts differentially on male versus female functions (e.g. Charlesworth et al., 1987; Queller, 1987; see also Charnov, 1979). Given either definition, the rewarding floral displays of animal-pollinated plants are widely acknowledged as products of sexual selection (reviewed by Andersson, 1994). Still, the analogy between floral traits and the exaggerated signals and weaponry of animals is problematic (Grant, 1995). Plants do not compete directly for mates; they compete for the attention of foraging pollinators. More importantly, current concepts of sexual selection in plants do not necessarily imply the existence of exaggerated secondary sexual characters.

Here, we recognize a type of frequency-dependent selection as the fundamental feature of secondary sexual characters. Frequency dependence typically arises in competitive interactions, including competition over mates, because the success of a particular trait value should depend on the average trait value of competitors. In particular, when individuals with a larger or more elaborate trait than average have a competitive advantage (herein ‘asymmetric competition’, after Matsuda & Abrams, 1994; Kisdi, 1999), selection may favor escalating investment in competing traits despite rising viability costs to the average individual (e.g., Abrams & Matsuda, 1994 and refs therein) and despite, in some cases, a decline in the mean fecundity of the population (e.g. Price et al., 1993; Matsuda & Abrams, 1994; Day & Taylor, 1996). This lack of an absolute optimum in asymmetric competition accounts for the ‘extravagant’ quality of signals and weaponry used in competition over mates. Likewise, the same type of selection can explain exaggerated traits used in competitive contests in general (West-Eberhard, 1983; Zahavi, 1991).

Many existing models of investment in pollinator attraction necessarily involve frequency dependence by treating attractive traits as a component of sex allocation (e.g. Charnov and Bull, 1986, Charlesworth & Charlesworth, 1987; Lloyd, 1988; Morgan, 1992). The fact that diploid individuals have exactly one mother and one father means that individuals in each generation compete for fixed genetic profits available through male and female functions (e.g., Charnov, 1982; Frank, 1990). The fraction of these profits that a plant obtains (i.e. reproductive success) depends on its competitive ability through a particular gender function (somehow related to investment in attraction) relative to the competitive ability of conspecifics. But beyond this Fisherian frequency-dependent competition, which exists in all models of sex allocation, it is
sometimes appropriate to include an additional component of frequency dependence in the measure of competitive ability (Maynard Smith, 1980; Frank, 1990). One such context occurs when a plant’s ability to attract pollinators depends on the traits of its competitors, but to our knowledge, no model of sex allocation in plants has considered this problem. Furthermore, most sex allocation models only consider how a fixed pool of reproductive resources should be divided among various components (e.g. pollen, seeds, and pollinator attraction; Charlesworth & Morgan, 1991). Hence, most current models do not include a viability cost for further investment in any one component; instead, the equilibrium allocation depends only on the ratio of marginal benefits for additional allocation to each component (Lloyd, 1988).

We introduce two models to explore the influence of asymmetric competition for pollinator visits on the adaptive dynamics of rewarding plant traits. First, a simple model of sex allocation in cosexual plants is extended by adding an extra component of frequency dependence that arises from the competition for pollinators. Second, we model investment in pollinator attraction as a life-history strategy, in which further investment increases the total size of the reproductive resource pool but also incurs a viability cost (shared by male and female functions of a cosexual plant). This model illustrates that asymmetric competition among animal-pollinated plants can favour traits that are strictly analogous to the secondary sexual traits of animals. We also show that exaggerated traits can evolve despite strongly diminishing returns for attractiveness which may otherwise seem to impose an absolute limit on rewarding traits (Klinkhamer & de Jong, 1993). Finally, in the Appendix, competitive and facilitative interactions among related plants are considered. Overall, the intention is to highlight the conditions in which exaggerated floral traits might exist in nature.

2.1.1 Pollinator attraction: background

In the models below, I assume (after Lloyd, 1988; Morgan, 1992) that pollinator visits are to some extent limiting the male and female functions of a cosexual plant (pollen export and import, respectively). In this case, any trait that increases attractiveness (the frequency of pollinator visits to a plant) provides an initial advantage. The same rewarding traits that increase a plant’s attractiveness, however, also tend to increase the number of flowers probed per visit (herein ‘residence’) (Klinkhamer and deJong, 1993; reviewed by Ohashi & Yahara, 2001). When pollinators probe many flowers in succession, large amounts of pollen may be
removed and lost to self-pollination among a plant’s own flowers (geitonogamy). Geitonogamy can reduce the amount of pollen that would have otherwise been exported (pollen discounting) and, in self-compatible species, can reduce the number of ovules available for outcrossing (ovule discounting) (reviewed by Harder et al., 2004). Furthermore, the benefit of increasing pollen removal per visit may diminish because of local mate competition resulting from limited pollen dispersal within the population (Lloyd, 1980). On the whole, as residence increases, the benefit of further attractiveness is expected to diminish and eventually decline (the plant’s ‘dilemma’; Klinkhamer & de Jong, 1993; Klinkhamer et al., 1994; Finer & Morgan, 2003).

Given these costs and benefits of pollinator attraction, how much should plants invest in rewarding phenotypes? The solution can be complicated by frequency dependence in two ways (Cresswell & Galen, 1991):

- Pollinators’ average rate of visitation to plants with a particular reward value may depend on the mean value in the population. Pollinators consistently visit plants with large floral displays (many open flowers) more frequently than smaller conspecifics (reviewed by Ohashi & Yahara, 2001), and bumble bees in particular are able to locate and preferentially return to plants with augmented rates of nectar production (e.g., Thomson, 1988; Cartar, 2004; but see Klinkhamer & van der Lugt, 2004). Nevertheless, it is often unclear whether the visitation rate is some function of a plant’s absolute reward value or if it also depends on the average phenotype of competitors. Both cases are considered here.

- Pollinators’ mean residence on a plant with a particular reward value may depend on the average value in the population. The marginal value theorem (Charnov, 1976) predicts that a forager should leave a patch when its energy intake rate falls to that of the habitat on average. Consistent with this prediction, Cibula & Zimmerman (1987) found that bumble bees’ residence initially increases on nectar enriched plants but eventually converges to pre-enrichment levels when all conspecifics in the habitat are similarly enriched. Limited evidence also suggests that pollinator residence may increase with relative display size more so than absolute display size (Harder & Barrett, 1995). In the second model below, we consider both cases.
2.2 Models

We use a standard evolutionary invasion analysis (Otto and Day, 2006) to ask whether rare variant traits can invade a resident population that is fixed at a particular mean trait value. The analysis assumes that the population reaches a new fixed mean trait value and equilibrium population size before the next variant trait invades. The dynamics of invading traits are approximated by the fitness gradient (slope of the fitness of an invading variant trait with respect to the trait value) evaluated where the variant equals the resident population mean (see Abrams, 2001). Hence, we only consider variants that are very close to the resident mean value, consistent with a genetic system in which many loci make small additive contributions to a quantitative trait (Abrams et al., 1993). The trait of interest is a continuous measure of investment in rewards, assumed to represent either nectar production or the number of rewarding flowers displayed for pollinator attraction. In both models, we assume that plants do not vary in resource state. It is also assumed that pollen and seeds disperse completely, meaning that interacting plants are unrelated; in the appendix, a simple extension to related plants (with correlated trait values) is presented.

2.2.1 Model 1: Sex allocation with asymmetric competition

Models of hermaphroditic sex allocation that include investment in pollinator attraction (e.g. Charnov & Bull, 1986; Morgan, 1992; deJong and Klinkhamer, 2005) can be easily extended to include asymmetrical competition for pollinator visits. Suppose that plants have a strategy to allocate reproductive resources among three components: pollen \( r \), seeds \( f \), and pollinator attraction \( x \), where \( r + f + x = 1 \). The benefit of further allocation to either pollen or seeds is typically expected to diminish and is usually specified by a power function, \( k^a \), where \( 0 < a < 1 \). A variant strategy’s competitive ability through male \( M \) or female \( F \) components will be a function of its allocation to pollen \( r' \) or seeds \( f' \) multiplied by the success of pollen export or pollen import, respectively, due to pollinator attraction:

\[
M(r', x', x) = r'^a (x' - zx)^a 
\]  
(1 a)

\[
F(f', x', x) = f'^b (x' - zx)^b ,
\]  
(1 b)
where $x' - z x$ measures the attractiveness of the variant allocation to rewards, $x'$, in a population with mean investment $x$. The influence of $x$ on the attractiveness of the invading trait value is controlled by the parameter $z$, and throughout the paper, we assume $-1 \leq z < 1$. Because our methods assume only small deviations in $x'$ from the resident, the value of $x' - z x$ will be positive as long as $z$ is sufficiently smaller than 1. If $z = 0$, pollinator visits increase with the absolute value of $x'$, but when $0 < z < 1$, variants with a relatively large trait value gain an extra advantage in attracting pollinators. Diminishing returns for further attractiveness ($0 < \sigma, \phi < 1$) are usually expected, here due to the consequences associated with longer residence (e.g., de Jong & Klinkhamer, 2005). Differences in the exponents $\sigma$ and $\phi$ allow the benefit of further visitation to differ between male and female function.

The fitness, $\lambda$, of an invading trait (or strategy controlling multiple traits) is the per-capita number of individuals contributed to the next generation by individuals with the variant trait (strategy). In models of sex allocation, invasion fitness is given by the variant’s competitive ability through male and female functions relative to the average competitive ability of individuals in a monomorphic resident population (Shaw & Mohler, 1956; Charnov, 1982):

$$\lambda(r', f', x', r, f, x) = \frac{1}{2} \left\{ \frac{M(r', x', x)}{M(r, x, x)} + \frac{F(f', x', x)}{F(f, x, x)} \right\}, \quad (2)$$

where the $\frac{1}{2}$ describes transmission of the trait value through male and female gametes. Notice that $\lambda = 1$ in the resident population, so only a variant trait combination giving $\lambda > 1$ can invade. Because $r + x + f = 1$, we can rewrite eq. 2 as a two-variable model in its full form (combining eqs. 2 and 1) as

$$\lambda_1(f', x', f, x) = \frac{1}{2} \left\{ \left( \frac{1 - f' - x'}{1 - f - x} \right)^{\alpha} \left( \frac{x' - z x}{x - z x} \right)^{\sigma} + \left( \frac{f'}{f} \right)^{\beta} \left( \frac{x' - z x}{x - z x} \right)^{\phi} \right\}, \quad (3)$$
A candidate evolutionarily stable strategy (ESS) $f^*, x^*$ satisfies the first derivative conditions
\[
\frac{\partial f^*}{\partial f^*} = 0 \quad \text{and} \quad \frac{\partial x^*}{\partial x^*} = 0,
\]
which specify the local trait values at which the fitness gradient is zero:
\[
f^* = \frac{\beta - z\beta}{\alpha + \beta + \sigma + \phi - z\alpha - z\beta}, \quad \text{(4 a)}
\]
\[
x^* = \frac{\sigma + \phi}{\alpha + \beta + \sigma + \phi - z\alpha - z\beta}, \quad \text{(4 b)}
\]
and by $1-f^* - x^*$,
\[
r^* = \frac{\alpha - z\alpha}{\alpha + \beta + \sigma + \phi - z\alpha - z\beta}. \quad \text{(4 c)}
\]

These equilibria will always be positive given that the exponents in eq. 3 are positive and $-1 < z < 1$. We have checked that the equilibrium $f^*, x^*$ is also convergence stable under these conditions, meaning that selection drives populations with mean values near $x^*$ even closer to the equilibrium (e.g., Christiansen, 1991). Finally, the equilibrium strategy must meet the ESS conditions
\[
\frac{\partial^2 \lambda^*}{\partial f^*} \leq 0 \quad \text{and} \quad \frac{\partial^2 \lambda^*}{\partial x^*} \leq 0,
\]
which ensure that the candidate ESS is a maximum on the fitness surface. We have found that these conditions do not depend on $z$, but otherwise, a neat analytical expression does not exist (see also Morgan, 1992). Under most reasonable parameter values, however (when the exponents in eq. 3 are less than 1 and not extremely different in magnitude), $f^*, x^*$ is in fact a fitness maximum (or local ESS).

Overall, Model 1 shows that with asymmetric competition for pollinator visits in the form $(x' - zx)$, the relative allocation to pollen, seeds, and pollinator attraction at the ESS is
\[
r^*: f^*: x^* = \alpha - z\alpha : \beta - z\beta : \sigma + \phi.
\]
When $z = 0$ (visits depend only on the absolute value of $x'$), we recover the usual result (e.g. Lloyd, 1988), in which relative allocations at the ESS are determined only by the ratio of exponents in the invasion fitness function: $\alpha : \beta : \sigma + \phi$. 
2.2.2 Model 2: Investment in pollinator attraction with viability costs

Here, we consider only the dynamics of $x'$, a variant investment in pollinator rewards that can take any positive value. We assume that a plant’s investment in primary male and female components (pollen and seeds) remains constant, so that the total size of the reproductive resource pool grows only by investing in rewarding traits. Due to energetic trade-offs, a plant is expected to incur an increasing viability cost with increasing $x'$. In addition to this energetic cost, there is a cost associated with excess pollen removal and geitonogamy, resulting from longer pollinator residence. Treating these effects as a cost seems realistic, given that fitness returns may not only diminish but also become negative with increasing pollen removal and geitonogamous selfing (see Klinkhamer et al., 1994; Finer & Morgan, 2005).

As in Model 1, the fitness of a variant trait $x'$ in a population with mean value $x$ is proportional to the sum of its relative competitive abilities through male and female components:

$$\lambda(x', x) = \frac{1}{2} \left( \frac{M(x', x)}{M(x, x)} + \frac{F(x', x)}{F(x, x)} \right).$$

(5)

But the competitive ability of $x'$ through male and female function now depends on its success in asymmetric competition for pollinator visits (the benefit) and the costs, $V$, associated with energetic expenses and longer residence (with benefits and costs combining multiplicatively, after Morgan, 1994):

$$M(x', x) = (x' - zx)\sigma V$$

(6a)

$$F(x', x) = (x' - zx)\phi V.$$

(6b)

The parameters $\sigma$ and $\phi$ describe how the benefit to male and female function, respectively, scales with further pollinator visitation, and again, we assume that $0 < \sigma, \phi < 1$ (visitation rate usually increases at a decelerating rate with further investment in attraction [Iwasa et al., 1995]). For simplicity, we assume that all costs affect male and female components equally. However,
we analyze the ESS investment in rewards under two different forms of \( V \), depending on whether pollinator residence increases with the absolute or the relative value of \( x' \).

### 2.2.2.1 Model 2a. Pollinator residence increases with absolute value of \( x' \)

The costs of investing in pollinator rewards are described by a negative exponential function of \( x' \). In this first case, the cost is a function of the absolute value of \( x' \) only:

\[
V(x') = e^{-(cx' - cx')},
\]

(7)

where \( c \) and \( g \) control the magnitude of the cost due to energetic expenditure and due to increasing residence, respectively (\( 0 < c, g < 1 \)). Combining eqs. 7, 5, and 6, the invasion fitness of \( x' \) in a population with mean value \( x \) is

\[
\lambda_{2a}(x', x) = \frac{1}{2} \left[ \left( \frac{x' - x}{x - x} e^{-(cx' - gx')} \right)^{\sigma} e^{(cx' - gx')} + \left( \frac{x' - x}{x - x} \right)^{\phi} e^{(cx' - gx')} \right]
\]

(8)

(see Fig. 2.1 a, c). The fitness gradient \( \frac{\partial \lambda_{2a}}{\partial x'} \bigg|_{x=x^*} \) is zero when

\[
x^* = \frac{\sigma + \phi}{2(c + g)(1 - z)},
\]

(9)

which is always positive and convergence stable when \( z < 1 \). Finally, the second derivative of the fitness gradient is negative if

\[
\frac{(c + g)^2(\sigma^2 + \phi^2 - 2\sigma - 2\phi - 2\sigma\phi)}{(\sigma + \phi)^2} < 0,
\]

(10)

which is always true when \( 0 < \sigma, \phi < 1 \) (meaning that \( x^* \) is a local ESS). From eq. 9, it is clear that the ESS investment in rewarding traits increases (holding all else constant) whenever \( \sigma \) and/or \( \phi \) increase (the benefit of pollinator visits increases more linearly) and whenever the costs, \( c \) and/or \( g \), decrease. Furthermore, when attractiveness to pollinators depends only on the absolute value of rewarding traits (\( z = 0 \), the ESS is always smaller than when plants compete
for pollinator visits in an asymmetric competition \( (z > 0; \text{Fig. 2.2}) \). On the other hand, facilitation of pollinator visits \( (z < 0) \) reduces the ESS investment in rewards.

### 2.2.2.2 Model 2b. Pollinator residence increases with \( x'/x \)

Here, the cost is a function of \( x' \) and \( x \):

\[
V(x', x) = e^{(-cx' - gx'x)}.
\]

Combining eqs. 11, 5, and 6, the invasion fitness of \( x' \) is

\[
\lambda_{2b}(x', x) = \frac{1}{2} \left\{ \left( \frac{x' - zx}{x - zx} \right)^{\sigma} \frac{e^{(-cx' - gx'x)}}{e^{(-cx - g)}} + \left( \frac{x' - zx}{x - zx} \right)^{\phi} \frac{e^{(-cx' - gx'x)}}{e^{(-cx - g)}} \right\}
\]

(see Fig. 2.1 b, d), and in this case, the candidate ESS is

\[
x^* = \frac{\sigma + \phi + 2g(z - 1)}{2c(1 - z)},
\]

which is always positive when \(-1 \leq z < 1\) and \(2g(z-1) < \sigma + \phi\). Again, positive values of the equilibrium are always convergence stable, and \( x^* \) is a local ESS when

\[
\frac{c^2(\sigma^2 + \phi^2 - 2\sigma - 2\phi - 2\sigma\phi)}{(\sigma + \phi + 2g(z - 2g)^2) < 0},
\]

which, as above, is satisfied under the assumption \( 0 < \sigma, \phi < 1 \). As in model 2a, increasing \( z \) also increases \( x^* \), but the ESS is exaggerated even further when pollinator residence increases with \( x'/x \) rather than \( x' \) (Fig. 2.2).

### 2.3 Discussion

Some authors have suggested that the sedentary nature of flowering plants makes the competition for pollinators unlike the direct competition for mates among animals (e.g., Grant, 1995; Murphy, 1998). Here, we suggest that both types of competition imply the same
fundamental type of selection. Asymmetric competition for pollinators can shift sex allocation to rewarding traits at the expense of pollen and seeds (Model 1) or can exaggerate investment in rewards despite viability and geitonogamy costs (Model 2). Hence, our models seem to predict the same ‘extravagant’ quality of secondary sexual traits that prompted Darwin’s (1871) conception of sexual selection in animals. One unique exception is that in dioecious animal populations, selection for an exaggerated trait is typically biased towards a particular gender function. When this is true of a cosexual plant population, however (i.e., exponents $\sigma$ and $\phi$ differ), trait exaggeration is effectively constrained by the gender function with the more diminishing benefit (the smaller exponent; Morgan, 1994). As a result, exaggeration of rewarding traits can be greatest when the benefit of further attractiveness is large and unbiased towards male or female function ($\sigma$ and $\phi$ are both close to 1).

Considering the costs associated with longer pollinator residence (the dilemma of pollinator attraction), some have suggested that rewarding traits should be constrained to ‘intermediate optima’ (Klinkhamer & de Jong 1993). This predicted constraint on pollinator rewards, though not always clearly defined, seems to imply an absolute limit. We have shown here that, even when pollinator residence increases with the absolute investment in rewards (Model 2a), asymmetric competition for pollinator visits tips the balance of the dilemma towards accepting the costs of excess pollen removal and geitonogamy. Investment in attraction can be exaggerated even further if residence increases with only the relative investment in rewarding traits (Model 2b; motivated by nectar enrichment experiments of Cibula & Zimmerman, 1987). Such a scenario can arise if plants with relatively rapid nectar secretion attract more pollinator visits and also encourage longer residence (e.g. Cartar, 2004). When the competing trait is the number of rewarding flowers displayed, it is unclear whether Model 2a or 2b should apply. Although relatively large floral displays consistently increase both pollinator visitation rate and the length of residence (Ohashi & Yahara, 2001), future studies should consider how residence on a particular floral display size depends on the mean display size of competitors (see Chapter 2).

2.3.1 Does asymmetric competition for pollinator visits exist?

A critical test of the models in this paper is to ask whether the attractiveness of a plant depends only on its own trait value, or whether it also depends on the mean trait value of competing
con specifics. In particular, we assumed that the attractiveness of a particular reward value declines as the mean reward value of competitors increases. Few studies have manipulated a rewarding trait to measure the visitation response to both manipulated and non-manipulated plants in a population. In one relevant study, however, Thomson (1988) found that nectar enriched *Aralia hispida* plants received more visits than before the manipulation, while non-enriched plants received fewer visits than expected. This result is consistent with asymmetric competition, but it would also be important understand whether such changes in visitation can be maintained in the long term (given pollinator population dynamics, for example). Rather than perform such long-term experiments, perhaps the best way to study trait interactions is to measure the visitation rate to focal plants with particular trait values in different patches or populations that naturally vary in mean trait value (‘contextual analysis’, e.g., Heisler & Damuth, 1987). Studies of this kind would be particularly useful in mass flowering species or in other populations with seemingly extravagant rewards.

In contrast to Thomson’s study, Klinkhamer et al. (2001) found that plants with relatively high rates of nectar production can facilitate visits to less rewarding neighbours. In *Echium vulgare*, the rate of visitation to plants in mixed patches of low and high nectar producers increased relative to patches of uniformly low producers; once attracted to the mixed patch as a whole, however, pollinators apparently visited low and high quality plants indiscriminately. When neighbouring plants are unrelated, facilitative interactions generally reduce the ESS investment in rewarding traits in favour of allocation to pollen and seeds (Model 1) or greater viability (Model 2) relative to cases in which \( z = 0 \) (see also de Jong & Klinkhamer, 2005). The results of Klinkhamer et al. (2001; also Klinkhamer & van der Lugt, 2004) emphasize that asymmetric competition through nectar production is only likely in populations where site-faithful foragers can memorize and discriminate among individual plants that differ in quality (e.g. Thomson, 1988; Cartar, 2004). On the other hand, if the competing trait is the number of rewarding flowers displayed, then pollinators need only associate larger floral displays with greater reward.

### 2.3.2 Evidence of a viability cost for rewarding traits

An important difference between our models is whether additional investment in pollinator rewards is at the expense of reproductive resources only (Model 1) or plant viability in general (Model 2). Consistent with sex allocation models, some evidence suggests that allocation
tradeoffs exist among floral reproductive organs (e.g. Mazer et al., 1999; reviewed by Charlesworth and Morgan 1991). But evidence for a tradeoff between pollinator attraction and other life-history components in cosexual plants is scarce. To our knowledge, only Pyke’s (1991) manipulation of nectar production rate in *Blandfordia nobilis* is fully comparable to experimental evidence of viability costs in the secondary sexual traits of animals (e.g. Moller, 1989). In Pyke’s study, artificially elevated rates of nectar production caused a reduction in seed set, but no detectable effect on plant size or belowground biomass (consistent with Model 1 and other sex allocation models). Viability costs of attractive traits have been detected in the males of dioecious plant species, however: *Silene latifolia* males that are selected to produce many flowers also grow smaller and experience physiological costs (Delph et al., 2005); and in a population of *Leucadendron xanthoconus*, males with the largest, most attractive floral displays also had the highest probability of dying (Bond & Maze, 1999).

Others have suggested that rewarding traits, especially nectar production, involve little energetic cost (e.g. Harder and Barrett, 1992). If this were true, our Model 2 predicts that competition for pollinator visits can exaggerate those traits to such large values that their total production cost becomes appreciable. Comparative analyses suggest that some species do in fact produce a surprisingly large amount of nectar (e.g., Harder and Barrett, 1992). In these species, it would be useful to investigate the viability costs of nectar production and the existence of asymmetric competition, which might be mediated by nectar production.

To summarize, a unified theory of sexual selection in animals and plants has so far relied on broad definitions of sexual selection, which do not necessarily imply the existence of exaggerated secondary sexual traits. We have shown that, with asymmetric competition for pollinator visits among conspecifics, the rewarding traits of animal-pollinated plants can be placed in the same class as the exaggerated secondary sexual traits of animals. Our analysis is a first step, and it is possible that other conditions (e.g., interactions among related plants; see Appendix) may alter the outcome. Nevertheless, few empirical studies, including those guided by sexual selection theory, have appreciated the potential significance of asymmetric competition among plants. This focus could fill a significant gap in our understanding of floral adaptations, however, and will be a valuable contribution to the formal investigation of sexual selection in plants.
2.4 Literature Cited


2.5 Appendix: Investment in pollinator attraction in a simple patch-structured population

Here, I consider a simple extension to Model 2b in the main text, where, due to clonal growth, the traits of ‘individuals’ (ramets) within a local patch may be positively correlated. All pollen and seeds are assumed to disperse from the patch, so there is no local competition among the gametes or offspring of related individuals. The analysis is based on a simplified version of the neigbour-modulated fitness approach (e.g., Taylor & Frank, 1996; Frank, 1998).

Consider a focal individual with investment in rewards $x'$ in a local patch with a mean trait value $y$ and a resident population with mean trait value $x$. Assuming that pollinator visitation and patch residence depends on the focal individual’s trait and the mean trait value in its patch, the fitness ($w$) of the focal individual will be

$$w(x', y, x) = \frac{1}{2} \left[ \left( \frac{x' - zy}{x - zx} \right)^\sigma \frac{e^{(c - gx'/(y))}}{e^{(c - g)}} + \left( \frac{x' - zy}{x - zx} \right)^\phi \frac{e^{(c - gx'/(y))}}{e^{(c - g)}} \right]$$

(A1)

(cf. eq. 12 in the main text). Using the chain rule, the marginal change in fitness with respect to the focal individual’s trait value is

$$\frac{dw}{dx'} = \frac{\partial w}{\partial x'} + \frac{\partial w}{\partial y} \frac{dy}{dx'}$$

$$= \left( \frac{-2g(z-1)+\phi + \sigma + 2c(z-1)x^*}{2(z-1)x^*} \right) + \left( \frac{2g(z-1)+(\phi + \sigma)}{2(z-1)x^*} \right) S'$$

(A2)

where all partial derivatives are evaluated at $x' = y = x = x^*$. The phenotypic derivative $dy / dx'$ is interpreted as a regression coefficient ($S$) that measures the correlation between the focal individual’s trait and the mean trait of its patchmates. The first term in eq. A2 is the direct effect of a marginal change in $x'$ on the focal individual’s fitness, and the second term measures the fitness effect from its social partners. By solving $\frac{dw}{dx'} = 0$, the candidate ESS is

$$x^* = \frac{2g(S-1)(z-1) + (Sz-1)(\phi + \sigma)}{2c(z-1)}.$$
Figure 2.A.1 illustrates the change in $x^*$ with increasing $z$ at varying levels of the correlation among patchmates ($S$). Under asymmetric competition ($z > 0$), the evolutionarily stable investment in rewards is expected to decline with increasing $S$. In this case, neighbours with correlated traits should cooperate by restraining their allocation to costly competition. In the case of facilitation ($z < 0$), on the other hand, $x^*$ is expected to increase with increasing $S$. Here, correlated neighbours should cooperate by facilitating pollinator visits (a ‘public good’) to one another.
2.6 Figures

Figure 2.1

The fitness $\lambda$ of a rare variant trait value $x'$ in a resident population with mean value $x$, from models 2a [(a) and (c)] and 2b [(b) and (d)]. The value of $\xi$ represents the value of both exponents $\sigma$ and $\phi$ in equations 8 and 12, which control how fitness benefits increase with further pollinator attraction. With all other parameters held constant ($z = 0.2, c = 0.1, g = 0.2$), the average peak of the fitness surface shifts to higher values in model 2b than in 2a [compare (a) to (b) and (c) to (d)]. Also notice that in model 2b, the fitness peaks shift with an increase in $x$ more so than in model 2a [compare (a) to (c) and (b) to (d)].
Figure 2.2

The evolutionarily stable investment in rewards \((x^*)\) as a function of the strength of asymmetric competition for pollinators \((z)\), as predicted by Models 2a and 2b from the main text. All other parameters are held constant \((\sigma = 0.9, \phi = 0.9, c = 0.1, g = 0.4)\).
Figure 2.A.1

The evolutionarily stable investment in rewards \( (x^*) \) as a function of the strength of asymmetric competition for pollinators \( (z) \), as predicted by a patch-structured version of Model 2b from the main text. The parameter \( S \) measures the correlation between a focal variant individual’s trait value and the average trait value of its patchmates. All other parameters are held constant \( (\sigma = 0.9, \phi = 0.9, c = 0.1, g = 0.1) \).
Chapter 3
Habitat assessment ability of bumble bees implies frequency dependent selection on floral rewards and display size

3 Abstract

Foraging pollinators could visit hundreds of flowers in succession on mass flowering plants, yet they often visit only a small number, potentially saving the plant much self-pollination among its own flowers (geitonogamy). This study tests the hypothesis that bumble bee (Bombus impatiens) residence on a particular plant depends on an assessment of that plant’s reward value relative to the overall quality experienced in the habitat. In a controlled environment, naive bees were given experience in a particular habitat (all plants having equal nectar quality or number of rewarding flowers), and we tested whether they learn about and adaptively exploit a new habitat type. Bees’ residence on a plant (number of flowers probed per visit) was eventually invariant to a doubling of absolute nectar quality and increased only slightly with a doubling of absolute flower number in the habitat. These results help to explain why pollinators are quick to leave highly rewarding plants and suggest that the fitness of rewarding plant traits will often be frequency dependent. One important implication is that geitonogamy may be a less significant constraint on the evolution of rewarding traits than generally supposed.

3.1 Introduction

The type of selection acting on the reproductive traits of animal-pollinated plants will often depend on the details of pollinator foraging behaviour. Behaviour plays a major role, for instance, in the conflict between attracting many visitors to a plant but also minimizing self-pollination among its flowers (geitonogamy)—a central theme in the evolutionary ecology of

---

plant reproduction (recent reviews by Barrett 2003; Harder et al. 2004; deJong & Klinkhamer 2005). A conflict can arise if the same traits that increase a plant’s attractiveness to pollinators also increase the number of flowers probed per visit (‘residence’ on the plant) (Charlesworth & Charlesworth 1987; Klinkhamer & deJong 1993). Plants that display more rewarding flowers than competing conspecifics, for example, attract more pollinators but can also encourage longer residence and more geitonogamy (e.g., Harder & Barrett 1995; Eckert 2000; Karron et al. 2004). Given these results and the expected costs of geitonogamy (e.g., pollen discounting, inbreeding depression in selfed seeds; see Harder et al. 2004), many authors have supposed that selection sets an upper limit on plants’ investment in rewarding traits (e.g., Wyatt 1980; Klinkhamer & deJong 1993). The existence of mass flowering displays can therefore seem contradictory (deJong et al. 1993; Harder et al. 2004), and by the same logic, so might the staggering cumulative nectar production of such displays (see Dupont et al. 2004).

This logic, however, implies that pollinator residence on a plant simply increases with the plant’s absolute reward value, and in some cases (e.g., deJong et al. 1992), models of reproductive investment have assumed that pollinators probe every flower per visit to a plant. In reality, pollinators of mass flowering species probe only a small proportion of flowers per visit (reviewed by Robertson 1992; Harder et al. 1994). Furthermore (and perhaps consistent with this observation), foraging theory suggests that pollinator residence on a plant should depend on the relative advantage of leaving to forage elsewhere. The marginal value theorem (MVT; Charnov 1976) suggests that if foraging gains decelerate with further residence on a plant (‘depression’; e.g., due to increasing risk of revisiting flowers; Ohashi & Yahara 2001), then pollinators would do best by leaving when the rate of reward intake falls to the maximum possible rate in the habitat (assuming complete knowledge of the habitat). Hence, on a plant with a particular reward value, optimal foragers should probe fewer flowers (and transfer less self pollen) as the reward quality of the habitat increases. From the plant’s perspective, this means that the fitness of rewarding traits may often depend on the frequency of other rewarding traits in the population or community (Pyke 1980; Creswell & Galen 1991; Biernaskie & Elle 2007).

The evolutionary dynamics of rewarding traits can be altered significantly when considering a geitonogamy cost that is frequency dependent. If, for instance, pollinator residence increases only with a plant’s relative (but not absolute) reward value (Biernaskie & Elle 2007), then a
relatively rewarding mutant can invade a resident population when the benefit of attracting pollinators initially outweighs the geitonogamy cost. But once the mutant trait value becomes common, pollinators probe the same number of flowers/plant as in the former resident population, thus ‘re-setting’ the cost and potentially favouring even further investment in rewarding traits. Still, the extent to which pollinator residence on a plant actually varies with the relative and absolute value of traits such as nectar quality and floral display size has not been studied in detail.

If pollinators are to behave according to the MVT, they must first use a learning rule that assesses the quality of their current habitat and adjusts their departure threshold toward the optimal value (cf. McNamara & Houston 1985). Here, we report a controlled experiment that tests whether bumble bees (*Bombus impatiens*) with prior experience in a particular habitat (all plants having equal nectar quality or number of rewarding flowers) can learn about and adaptively exploit a new habitat type (with different nectar quality or display size). Evidence from field experiments suggests that bees may in fact adjust plant residence to the mean quality of a population (Hodges 1985; Cibula & Zimmerman 1987; Harder & Barrett 1995), but in these cases it was impossible to fully control the prior experience of individual foragers or to independently vary display size and floral rewards. Furthermore, the current study isolates the effects of prior information and learning by comparing the behaviour of naive and experienced bees in the same habitat type. If bees use the proposed learning rule, we predict the following:

- Bees should initially probe more flowers on plants that are nectar rich relative to prior experience (when ‘rich’ plants are perceived as rare) but eventually treat both habitat types equally. This follows the MVT when all patches are of a single type: varying the absolute resource density of patches alone has no effect on optimal residence time (e.g., Charnov and Parker, 1995).

- Bees should initially probe more flowers on plants with relatively large floral displays (all else equal) but this should change with further experience in a large-display habitat. The risk of revisiting flowers likely increases more slowly on large displays than on smaller ones (Ohashi & Yahara 2002), meaning that bees can probe more flowers at each plant (and in this sense, large-display habitats are ‘richer’). But residence should be longer when large displays are perceived as rare than when perceived as common.
3.2 Materials and Methods

3.2.1 Bees, artificial plants, and general methods

Worker bumble bees (*B. impatiens* Cresson; supplied by Biobest Biological Systems, Canada) were trained to collect sucrose solution (herein ‘nectar’) from artificial flowers within a large screened enclosure (3.4 x 2.0 x 4.6 m$^3$). Bees entered the enclosure through a gated tunnel connected to their nest box. The ‘flowers’ were clear, 1.5 mL polypropylene microcentrifuge tubes with the cap removed and a circular collar (6.5 cm diameter) of blue cardboard fixed around the tube’s entrance. Flowers were grouped into inflorescences (herein ‘plants’) along a green Styrofoam tube (length = 75 cm; circumference = 6.5 cm) in two staggered columns, with flower openings spaced 5.5 cm apart. Each plant could hold up to 16 flowers (i.e., 2 columns of 8 flowers). Bees were initially trained to collect nectar from four flowers arranged haphazardly on a single plant that rested horizontally in the enclosure. When tests were not in progress, these flowers were filled as necessary to provide the colony with nectar, and this was the only foraging experience that bees had before an experimental trial. The concentration of floral nectar in the training flowers depended on the experiment being performed (details below).

Experimental trials consisted of two phases, during which a single marked bee made numerous consecutive foraging bouts to the enclosure. During all trials, 12 identical plants were presented in a linear array of two staggered columns, with plants spaced 60 cm apart. All flowers contained a 3 μL nectar reward, placed near the bottom of the tube. Depleted flowers were recorded and manually replenished after a bee’s departure from a plant; hence, the only empty flowers encountered were those revisited during a single plant visit (bees completely emptied all probed flowers). The first phase ended when the bee reached a specified number of total plant visits (excluding the last visit of every bout) and later returned to the colony to unload its crop. Before the next foraging bout, we replaced all plants with a new set for phase two. The flowers on these new plants were always made with washed centrifuge tubes and fresh cardboard collars. Throughout all experimental trials, we recorded the number of flowers probed/plant (including revisits) before departure to visit another plant in the array. Individual bees were tested only once and then removed from the colony.
3.2.2 Experiment 1: manipulation of nectar quality

This experiment used 16-flowered plants and two treatment levels that individual bees experienced in one of two orders: 20% (w/w) sucrose solution in all flowers (‘nectar poor’ habitat) for phase one and 40% sucrose in all flowers (‘nectar rich’ habitat) for phase two, or vice versa. Phase one lasted for 50 plant visits, and phase two lasted for at least another 50 visits (we sometimes recorded two extra foraging bouts to increase the likelihood of observing learned behaviour). The goal of the nectar manipulation was to alter the rate of energy gain per flower without varying any other aspects of foraging economics. Sucrose concentration (as opposed to volume, for instance) was manipulated because the ingestion rate of a given volume of sucrose by *B. impatiens* is constant up to a concentration of 40-50% (Harder 1986). Hence, it is likely that our manipulation increased the energy gain per flower without affecting floral handling time.

The order of treatment levels (poor ⇒ rich, rich ⇒ poor) was assigned to individual bees from two different colonies. Bees that were assigned to begin with a particular treatment level were trained with that concentration of nectar. We could not randomly assign the order of treatment levels to individual bees within a colony, however, because it was only possible to train bees collectively. Instead, the first four bees tested from a colony were given one order (poor ⇒ rich, trained on 20% sucrose only) and the next four bees were given the opposite order (rich ⇒ poor, trained on 40% sucrose). We did the same with eight bees from the second colony but reversed the order of treatment levels.

3.2.3 Experiment 2: manipulation of floral display size

This experiment also had two treatment levels, but individual bees experienced them in one of three orders: all plants having 8 flowers in phase one (‘small display’ habitat) and then 16 flowers in phase two (‘large display’ habitat); all plants having 16 flowers in phase one and 8 flowers in phase two; or a random spatial distribution of six 8-flowered plants and six 16-flowered plants in both phases. The randomly mixed habitat was included as a type with intermediate frequency of large displays. All flowers in this experiment had 3 μL of 30% sucrose solution (and bees were trained only with 30% sucrose). Both phases lasted for 100 plant visits, but as above, phase two sometimes went slightly beyond 100 visits. In mixed
arrays, a new random distribution of display sizes was presented every five foraging bouts, and as always, plants were replaced in between phases.

Initially, the order of treatment levels (small ⇒ large, large ⇒ small, or mixed) was assigned to individual bees from two different colonies in a completely randomized design. This yielded six bees per treatment level order. Later, to increase our ability to detect differences between the small ⇒ large and mixed orders in particular, we randomly assigned four additional bees to each of these orders, using a third colony.

3.2.4 Data analysis

Because each bee experienced both treatment levels in a particular experiment, the factor ‘Bee’ (identifying individual subjects) is used as a random blocking factor throughout. The dynamics of bee behaviour as a function of experience in experimental trials is presented in leverage plots (JMP Start Statistics, SAS Institute, Inc.) of the mean number of flowers probed/plant at intervals of 5 (experiment 1) or 10 (experiment 2) plant visits. These plots are interpreted in the same way as simple regression plots, but they describe the marginal contribution of ‘Experience’ after adjusting for variation due to ‘Bee’. We summarize the relationships in leverage plots by calculating a linear regression slope for each bee and then estimating the mean slope and 95% confidence interval (CI) for a particular habitat type.

For each experiment, two split-plot ANOVA models were used to estimate the mean plant residence for particular order-treatment level combinations. All models included the factors ‘Order of treatment levels’ (acting at the level of individual bees) and ‘Treatment level’ (acting within each bee), their interaction, plus the random factor ‘Bee (nested within Order)’ to identify bees as the experimental units. Two response variables were associated with each bee: a mean plant residence for each treatment level. For the poor ⇒ rich and small ⇒ large orders (where we predicted a change in behaviour with experience in phase two), the response variable for ‘rich’ and ‘large’ was either a mean of the initial response in phase two (defined as the first 10 plant visits; see Model 1 in Table 3.1) or a mean of the ‘learned’ behaviour in phase two (defined as the last 25 visits in experiment 1 and last 50 visits in experiment 2; see Model 2 in Table 3.1). For all other treatment level orders, the mean residence was from all of phase one and all of phase 2, and in the mixed display size habitat, from the entire trial. From these
analyses, we report least squares (LS) means for every order-treatment level combination, and the differences between LS means, as estimated by Tukey tests. Models were estimated with the REML method, using JMP 5.01. Throughout the paper, means and 95% CIs are presented as: mean (lower limit, upper limit).

3.3 Results

Details of the ANOVA models are given in Table 3.1. Below, we focus on estimates of the effect sizes from those models and on the dynamics of bee behaviour with experience.

3.3.1 Experiment 1: manipulation of nectar quality

We first used this experiment to confirm that the probability of revisiting at least one flower on a 16-flowered plant increased with the total number of flowers probed on the plant (repeated measures logistic regression, using ‘Bee’ as the subject variable; G = 11.13; P < .0008; n = 16 bees). For example, during visits of 8, 10, and 14 flowers/plant, the probability of including at least one revisit was estimated as 0.12 (.06, .18), 0.27 (.19, .35), and 0.53 (.42, .64), respectively. This increasing probability of flower revisitation may be the primary reason that bees leave our artificial plants before probing every flower.

When habitat types were presented in the poor ⇒ rich order, the initial response of bees in phase two was to probe an average of 3.48 (2.50, 4.50) more flowers/plant than in phase one (Figs. 3.1a, 3.2a). Average plant residence gradually declined with experience in the rich habitat, however (Fig. 3.1a, top panel) and by the end of phase two, the mean residence did not differ significantly from the poor habitat in phase one [Fig. 3.2a; mean difference = 0.76 (-0.31, 1.85)]. Concurrent with this declining mean residence in the rich habitat was an increase in foraging efficiency: in the first ten plant visits, bees revisited an average proportion of 0.059 (0.048, 0.069; n = 8 bees) flowers/plant; by the last ten visits, a significantly smaller proportion of flower probes were revisits [0.019 (0.008, 0.029) flowers/plant].

The mean residence of naive bees in phase one of the rich ⇒ poor order did not vary with experience (Fig. 3.1a, lower panel), and was not detectably different than the learned response at the end of phase two of the poor ⇒ rich order (Fig. 3.2a). With prior experience in the rich
habitat, bees initially rejected all flowers in phase two (poor habitat), and in fact one bee stopped foraging altogether. When the others resumed normal foraging, however (usually after ~5 mins. of searching the habitat), mean plant residence in the nectar poor habitat did not differ significantly from the rich habitat in phase one [Fig. 3.2a; mean difference = -0.46 (-1.50, 0.59)].

3.3.2 Experiment 2: manipulation of floral display size

Bees in the small ⇒ large order initially probed an average of 4.04 (2.84, 5.24) more flowers/plant on large displays than on the small displays in phase one (Figs. 3.1b, 3.2b). But as above, plant residence in the large display habitat gradually declined with experience in phase two (Fig. 3.1b, top panel), so that the mean difference between phase one (small displays) and the learned behaviour in phase two (large displays) was only 1.80 (0.59, 3.00) flowers probed/plant (Fig. 3.2b). The LS mean residence on large displays by the end of the small ⇒ large order was significantly less than the mean residence on large displays in the mixed display habitat [mean difference: 2.41 flowers/plant (0.64, 4.18); see Fig. 3.2b]. Incidentally, we note that bees in our mixed habitat encountered large and small plants completely randomly, rather than showing a preference for visiting large displays: bees visited large displays at a mean frequency of 50.6 % (48.3, 52.9; n = 10 bees). This was true even though bees had the ability to discriminate between our small and large display sizes in the experimental arrays (data not shown).

The mean residence of naive bees in phase one of the large ⇒ small order did not vary with experience, (Fig. 3.1b, lower panel), but the mean number was surprisingly large and significantly greater than the learned response of bees in phase two of the small ⇒ large order (Fig. 3.2b). Unfortunately, the average behaviour of naive bees in this large display habitat was skewed by two bees (out of a sample of six) that were particularly persistent on large displays. Notice, however, that all plants in experiment 1 also had 16 flowers and that bees in phase one of this experiment can also be considered as naive to the large display habitat. The LS mean number of flowers probed/plant by these naive bees (shown as ♦ in Fig. 3.2b) was similar to the learned response of bees in phase two of the small ⇒ large order.
3.4 Discussion

Our results demonstrate that bumble bee residence on a particular plant depends not only on the reward quality of that plant, but also on the expected foraging gain elsewhere, which is learned through experience. Bees initially stayed much longer on plants in a habitat that was more rewarding than their previous experience. But given further experience with rewarding types, the average plant residence was nearly invariant to absolute nectar quality and increased only slightly with the number of rewarding flowers/plant in the habitat. We discuss these results in relation to foraging theory below. Our second objective is to emphasize how the foraging behaviour observed here can translate into frequency dependent selection on floral rewards and display, which has been largely overlooked (but see discussions by Pyke 1980; Harder & Barrett 1996; Morgan 2000; Biernaskie & Elle 2007; Ferdy & Smithson 2002 consider frequency dependence in another context). Our study suggests that to fully appreciate the adaptive evolution of rewarding traits, it is necessary to consider how selection acts on a rare, relatively rewarding trait value but also how selection changes if that trait value were to become common.

3.4.1 Contribution to foraging theory

The experimental environment of the current study was very similar to the assumptions of the MVT in its simplest form (i.e., most habitats of a single patch type, constant travel time within and between patches, rewards available at every flower). The MVT specifies the optimal behaviour when foragers have complete information about such an environment. In reality, foragers might use simple rules to learn the environmental parameters and the appropriate patch residence time as specified by the MVT (McNamara & Houston 1985). Our study suggests that bumble bees can learn to adjust plant residence toward a seemingly adaptive value and that this rule depends on initial estimates of habitat quality (which we imposed experimentally). We did not directly measure whether bees maximized the long-term rate of reward intake (‘optimality’ per se). However, bees did stay longer on a relatively rewarding plant when the estimate of average intake rate should have been low, and mean residence gradually declined as experience was swamped with the more rewarding habitat type. This behaviour seems adaptive in the sense that the decline in mean residence was also associated improved foraging efficiency (fewer flower revisits). Furthermore, if we assume that the mean residence adopted by naive bees in the richer habitat types is adaptive, then it is important to note that most experienced bees in the richer habitats eventually converged to that same average behaviour.
The use of information from prior foraging experience and current sampling in a patch is an example of Bayesian updating (recently reviewed by McNamara et al. 2006; Valone 2006). Few studies have provided convincing evidence of Bayesian foraging, but our design is particularly valuable because it allowed us to manipulate prior information but hold current sampling experience fixed (as advocated by Valone 2006). Naive bees acted as a ‘no prior information’ control, and we found that their behaviour in the richer habitat types did not vary over time. The fact that bees with experience in a less rewarding habitat did alter their behaviour over time in the richer habitats is strong evidence that prior information and learning altered these bees’ decisions. Given that bees had to learn only a uniform habitat quality, these results provide the most basic evidence for Bayesian foraging (McNamara et al. 2006). A test of whether bumble bees can also learn the distribution of patch types in an environment (the more difficult task) will be presented in a forthcoming study.

Our experiments suggest that the plant residence behaviour of bees depends on the mean reward value in a monomorphic habitat. Real populations, on the other hand, are often characterized by wide variation around the mean (e.g., Biernaskie & Cartar 2004; Biernaskie & Elle 2005). Hence, it may be necessary to ask whether such variation hinders pollinators’ ability to learn the mean quality of a population and behave accordingly. Bees also seemed to maintain a long ‘memory window’ of at least 20 or more plants in our experimental conditions (compare with Cibula & Zimmerman 1987), yet it would be interesting to know how much of their recent experience pollinators remember when foraging in a more variable or structured population. Despite these caveats, our results may have important implications for selection on rewarding plant traits.

3.4.2 Implications for selection on floral display size

It is well known that, within populations, pollinator residence on a plant increases with floral display size at a decelerating rate (reviewed by Ohashi & Yahara 2001). This was evident in our mixed display habitat, where bee residence did not increase in proportion with the doubling of display size (see Fig. 3.2b). We also showed, however, that the increase in residence on large displays is even smaller when they are common in the habitat. The effect known within populations is unlikely to fully explain the sharp decline in the proportion of flowers probed/visit with increasing mean display size among species (see Harder et al. 2004).
Consider, for example, that bee pollinators probe only about 9-12 flowers, on average, in populations of *Hybanthus prunifolius* (mean display size of 226 flowers; Augspurger 1980) and *Echium wildpretii* (mean display size of ~1800 flowers; introduced honey bees probe ~35 flowers/plant; Dupont et al. 2004). Given our results, the fact that mass flowering plants typically compete with many other mass flowering individuals in the population (or community) may help to explain why pollinators are quick to leave such large displays. An additional hypothesis which deserves attention is that mass flowering plants may be visited so frequently that pollinators often find empty flowers, encouraging early departure (Harder et al. 2004). Yet in our experiments, where flowers were always rewarding (save revisits), it is unclear why both naive and experienced bees should visit so few flowers in our large display habitats. Revisits were generally rare (~3% of all flower visits), meaning that bees usually left a plant after probing a rewarding flower and suggesting that the perceived risk of revisitation may be a key determinant of plant departure (see also Ohashi and Yahara 2001, 2002).

For selection to favour further investment in floral display size, some benefit of relatively large displays must outweigh the potential cost of longer residence and increased geitonogamy. If a benefit exists, however, the behaviour observed in our study suggests that this initial cost should lessen as larger displays become common, potentially favouring even further investment in display size. One likely benefit of relatively large displays is that they attract more visitors (reviewed by Ohashi & Yahara 2001). Furthermore, it seems plausible that the attractiveness of a particular display size declines as the mean display size of the population increases, ensuring that relatively large displays have a competitive advantage (‘asymmetric competition’, cf. Biernaskie & Elle 2007). This form of competition for pollinator visits can favour escalating investment in display size, even if pollinator residence and geitonogamy increase with the mean display size of the population (Biernaskie & Elle 2007; see also Morgan 2000). In some species, mass flowering is in fact maintained despite the occurrence of significant geitonogamy costs (e.g., Eckert 2000; Dorken et al. 2002).

### 3.4.3 Implications for selection on floral rewards

We found that bees’ residence on a plant is eventually independent of the absolute quality of nectar per flower in a habitat, and we suggest that this result might extend to floral rewards in general (e.g., nectar volume, pollen availability: any trait that increases the per flower net intake
rate). In some cases, plants with a relatively rapid rate of nectar production can encourage frequent pollinator visitation to the plant and also promote longer residence per visit (e.g., Cartar 2004). When this is true, further investment in nectar production may be favoured if a benefit of attracting frequent visitors outweighs the potential geitonogamy cost associated with a relatively high rate of nectar production. Our results suggest that this initial cost may be effectively ‘reset’ once any particular investment in nectar production becomes common in the population. On the other hand, if pollinator visitation rate can eventually match plants’ rate of nectar secretion, then pollinators will not actually encounter larger nectar rewards per visit, and residence on a plant should not depend on the rate of nectar production (deJong & Klinkhamer 2005).

Finally, and in contrast to our emphasis so far, floral rewards may sometimes function to encourage pollinator residence on a plant when geitonogamy is not of great consequence. Encouraging longer residence may be adaptive if plants already have effective mechanisms to minimize geitonogamy (e.g., herkogmany and dichogamy; reviewed by Barrett 2003) or if pollinators are so scarce that the main problem for a plant is to attract at least some visitors (Iwasa et al. 1995; Harder et al. 2001). In the latter case, large floral displays may be used to attract pollinators to the plant, while floral rewards ensure that pollinators stay long enough to engage most of the plant’s pollen in dispersal and to import enough pollen to fertilize its ovules. In this context, geitonogamy can evolve as a by-product of selection for attraction and longer residence and, in some cases, may even act as a form of reproductive assurance (Lloyd 1992; Harder & Johnson 2005). The intriguing implication of our results is that further investment in floral rewards may be continually favoured (checked only by rising energetic costs or other constraints) because longer pollinator residence (a benefit, in this case) is a consequence of increasing the relative, but not absolute, value of rewards. Although previously overlooked, this would be another example of the type of selection that is strictly analogous to sexual selection in animal populations (see Biernaskie and Elle (2007)). Some plant species do in fact seem to produce an abundance of rewards (see Schemske, 1980; Harder & Barrett 1991; Dupont 2004), and it would be interesting to know if a frequency dependent competition accounts for this apparent extravagance.
3.5 Literature Cited


3.6 Figures

Figure 3.1

Leverage plots of the relation between the number of flowers probed per plant and experience in experiment 1 (a) or experiment 2 (b), where the dotted line separates phases one and two. A mean is given for each bee at each interval of experience. In (a), the mean regression slope across bees in phase two of the poor ⇒ rich order was -0.28 (95% CI: -0.41, -0.15; n = 8 bees), but otherwise, the mean slope was not significantly different than zero in any other habitat type. In (b), the mean regression slope in phase two of the small ⇒ large order was -0.25 (-0.42, -0.14; n = 10), but the mean slope was not significantly different than zero in any other habitat type.
Figure 3.2
Least squares mean number of flowers probed in the habitat types of experiment 1 (a) and experiment 2 (b). The mean initial response of bees in phase two of the poor $\Rightarrow$ rich and small $\Rightarrow$ large orders is denoted by ‘X’; otherwise, the mean value in phase two of these orders represents the learned behaviour. The mean values for the mixed order in (b) were taken over the entire trial (the mean difference in residence on large and small plants did not differ significantly between phase one and two). The mean shown as $\Diamond$ in (b) is of bees in experiment 1 that were also naive to the large floral display habitat (details in text). Error bars
give the 95% confidence interval, and means labeled with different letters are significantly
different by a Tukey’s HSD comparison.
3.7 Tables

Table 3.1

ANOVA results from experiment 1 [manipulation of nectar quality; (a)] and experiment 2 [manipulation of display size; (b)]. In all models, there were two response variables for each bee (a mean residence behaviour for each treatment level). For the poor ⇒ rich and small ⇒ large orders, Model 1 uses the mean initial behaviour in phase two (first 10 plant visits of phase two), and Model 2 uses the mean learned behaviour in phase two [last 25 visits for (a) and last 50 visits for (b)]. For all other orders, the mean residence is an average of all of phase one or all of phase two, or an average of the entire trial [for the mixed order in (b)]. Each model also included the term ‘Bee [Order]’, but tests were not performed on this random factor. Least squares mean values from the interaction plots of each model are presented in Figure 3.2.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) Nectar quality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>26.5</td>
<td>1.23</td>
</tr>
<tr>
<td>Treatment level</td>
<td>38.5 (&lt;.0001)</td>
<td>0.33 (0.57)</td>
</tr>
<tr>
<td>Order x Treatment level</td>
<td>65.2 (&lt;.0001)</td>
<td>5.15 (0.041)</td>
</tr>
<tr>
<td><strong>(b) Display size</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>0.70 (2, 23)</td>
<td>6.31 (2, 23)</td>
</tr>
<tr>
<td>Treatment level</td>
<td>240 (1, 23) (&lt;.0001)</td>
<td>154 (1, 23) (&lt;.0001)</td>
</tr>
<tr>
<td>Order x Treatment level</td>
<td>0.20 (2, 23)</td>
<td>7.64 (2, 23) (0.0029)</td>
</tr>
</tbody>
</table>
Chapter 4
Bumble bees learn to forage like Bayesians

4 Abstract

‘Bayesian’ foraging in patchy environments requires that foragers have information about the distribution of resources among patches (prior information), either set by natural selection or learned from past experience. We test the hypothesis that bumble bee foragers can rapidly learn prior information from past experience in two very different experimental environments. In the high variance environment (patches of low and high quality), stochastic optimality models predicted that finding rewards should sometimes sharply increase an optimal forager’s tendency to stay in a patch (an ‘incremental’ response), whereas in the uniform environment, finding rewards should always decrease the tendency to stay (a ‘decremental’ response). We use Cox regression models to show that, in a matter of hours, bees learned to match both predicted responses, resulting in a reward intake rate that averaged 80% of the predicted maximum. Following training in either environment, bees’ adaptive behavior carried over to a common test environment, thus confirming the influence of memorized prior information. Although Bayesian foraging by learning is often presumed of vertebrates, this study is the first to clearly isolate the adaptive use of a learned prior expectation. More generally, it highlights the remarkable adaptive plasticity of an important generalist pollinator and agent of selection.

4.1 Introduction

The history of foraging theory related to patch use shifts from considering when a forager should optimally leave a patch of resources (e.g., Charnov 1976) to how the forager should decide when to leave (reviewed by McNamara et al. 2006). This shift reflects an input of ecological realism to the patch-use problem; in particular, because food items are often

---

encountered stochastically and patches vary in quality, a forager must use incomplete information to decide when a patch should be abandoned. The optimal decision rule, which tells a forager how to infer future patch quality based on its current experience in a patch, can be calculated with Bayes theorem by an investigator who knows the distribution of food items among patches in the environment (the prior distribution; e.g., Oaten 1977; McNamara 1982; Green 2006). In order to approximate an optimal foraging rule, animals must similarly have some ‘knowledge’ of the prior distribution. They might learn prior information from their past experience; however, if the distribution of food items is stable over time, that information could instead be hardwired by natural selection, resulting in animals that behave as if they know the prior distribution for their environment (McNamara et al. 2006; Pierre and Green 2008).

Examples of patch-leaving behavior from both vertebrates (reviewed by Valone 2006) and invertebrates (see Pierre and Green 2008) can be interpreted as Bayesian-like, yet the extent to which prior knowledge of resource distributions is learned or hardwired remains unclear. Even the most impressive examples of Bayesian foraging by birds do not necessarily imply a special learning ability. In many cases, birds encountered an unmodified distribution of food items in nature (e.g., Alonso 1995) or an experimental distribution that might reflect the type of ancestral environment that the animals are already adapted to (e.g., the negative-binomial distribution studied by van Gils et al. (2003) or the empty/full distributions used by Lima (1984); see Ydenberg 1998). Furthermore, the birds used in many experimental studies were ‘trained’ to a particular food distribution for weeks. If learning did take place during this time, it is not clear how long it took for prior knowledge to be replaced by new information.

We hypothesized that an invertebrate forager, the bumble bee, should have the capacity for Bayesian-like foraging and the ability to rapidly learn prior information from very different types of environment. As generalists, bumble bees have an evolutionary history of encountering a diversity of nectar resources that can vary greatly among plants of the same species and among flowers within a single plant (Goulson 2003, p. 73 and references therein). Furthermore, the extent of variation among plants will differ among species that bees visit in a lifetime and may also fluctuate within a population as a result of daily patterns of reward depletion and renewal (e.g., Pleasants and Zimmerman 1979; Thomson 1988). It is exactly in this type of environment that natural selection would favor the ability to learn about the distribution of resources among patches.
When patches are consistently rewarding and of the same quality, bumble bees retain prior information about the mean quality of the environment and adjust their patch-leaving rules accordingly (Biernaskie and Gegear 2007). In contrast, the leaving rules that bees use on plants with variable rewards are often interpreted as fixed ‘rules of thumb’ (e.g. a threshold nectar volume or number of empty flowers: Pyke (1978); Hodges (1985)) or as rules that otherwise depend only on recent success in a patch (e.g., Kadmon and Shmida 1992). Yet given the variance in rewards that can be found within and among patches, bees that follow a fixed departure rule would often leave good patches too soon or stay too long in poor patches. They could avoid such errors by incorporating information about the distribution of rewards in their past experience.

The current study rigorously tests whether bumble bees can learn to adapt to novel environments by adjusting their patch-leaving rules in a Bayesian-like manner. The experiments involve training naive foragers (for hours, rather than weeks) in either a uniform environment (all patches having the same number of rewards) or a high variance (HV) environment (patches of either low or high quality) and then observing each forager in a common test environment. This design is unique in that it isolates the effect of memorized prior information on foraging behavior in the common environment and thus explicitly tests the central feature of Bayesian foraging via learning (McNamara et al. 2006; Valone 2006). We also take advantage of recent theoretical developments (Olsson and Holmgren 1998; Green 2006) to calculate the optimal patch-leaving rule in each training environment. Bumble bees are certainly not perfect Bayesian decision-makers, but optimality models allow us to compare their actual reward intake rates to the expected intake rate of an optimal forager. As a compliment to the optimal foraging models, we use a statistical model (extensions of Cox’s proportional hazards model; Therneau and Grambsch 2000) to estimate bees’ tendency to remain in a patch as a function of information gained by sampling the patch. In this way, the deterministic optimality predictions can be compared with a more realistic, proximate model of animal behavior (Wajnberg 2006; Pierre and Green 2008). We show a strong qualitative match between the two models, which, together with evidence from the test environment, strongly suggests the use of learned prior information in an adaptive, Bayesian-like way.
4.2 Materials and Methods

4.2.1 Experimental methods

Worker bumble bees (*Bombus impatiens* Cresson; supplied as eggs and larvae by Biobest Biological Systems, Canada) were trained to collect 30% sucrose solution (herein ‘nectar’) from artificial flowers within a large screened enclosure (3.4 x 2.0 x 4.6 m$^3$). Bees entered the enclosure through a gated tunnel connected to their nest box. The ‘flowers’ were clear, 1.5 mL polypropylene microcentrifuge tubes with the cap removed and a circular collar (6.5 cm diameter) of blue cardboard fixed around the tube’s entrance (Gegear and Laverty 2005). To train for initial flower recognition, 4-5 filled flowers were presented at a single location within the enclosure. Worker bees from the colony eventually found and foraged on these flowers to sustain the colony when experiments were not in progress. Active foragers were painted with a small, unique marking for identification.

For all training and tests, 12 flowers were grouped into rectangular patches (4 x 3), spaced 7.5 cm apart on pieces of rigid poster board. The ‘environment’ consisted of 10 patches, arranged in two rows of five and spaced 1 m apart. In total, 50 flowers in the environment each contained 5 μL of nectar, and the remaining 70 flowers contained 5 μL of water (details of the distribution of rewards are given below). During training and tests, only a single marked forager was allowed to make repeated foraging bouts to the environment, and while foraging, we controlled bees’ movements by covering all but two patches: the one being harvested and the (unvisited) adjacent patch. This ensured that bees’ inter-patch travel distance was consistent, that patches were never revisited, and that bees experienced the entire distribution of patch types that we intended. Individuals were trained on a particular distribution of rewards for at least 45 patch visits (usually 2-3 hours), consisting of multiple foraging bouts (length of bouts ranged from 4 to 10 patch visits). A bout ended when the bee returned to the colony to unload her harvest, and during this interruption, we refilled all rewarding flowers that had been visited in the environment (after checking that visited flowers were in fact emptied). Hence, a fully replenished environment was always available at the start of each new foraging bout.

We felt that 45 patch visits would be sufficient to test whether bees updated their pre-training prior distribution, presumably set mainly by evolutionary history, with new information about the distribution of rewards in the following training environments:
Uniform treatment. Each patch contained five rewarding flowers in a random spatial arrangement. After every second foraging bout, the position of patches in the environment was haphazardly shuffled, and after every 15 patch visits (or whenever the bout ended thereafter), a new set of patches with a new random arrangement of rewards was presented. We did this to minimize bees’ ability to memorize the location of rewarding flowers, but also to minimize the evaporation of rewards and the influence of floral scent marks.

High variance (HV) treatment. Half of the patches contained a single reward, and half contained nine rewards (all randomly distributed). The sequence of patch types in the environment was randomized. Patches were shuffled after every second bout, but in this case, patches were moved to a new randomized sequence. Also as above, a new set of patches was presented after every 15 patch visits.

Once the threshold of 45 training visits was exceeded (excluding the final patch of each bout), we replaced all patches with the test environment, and the test began on the next foraging bout. All patches in the test environment contained five rewarding flowers in a common spatial arrangement. Bees from both training environments experienced the common arrangement of rewards and the same sequence of patches within the test environment. The test lasted for only 10 patch visits so that the observed behavior in the test could reflect any learned information from prior training; in this way, we hoped to minimize the amount of new information that might be learned within the test itself. (This is particularly relevant for HV-trained (HV-T) bees, given that the HV environment was much different from the test).

Individual bees went through only one sequence of training (uniform or HV) and test. In total, 20 bees (10 of each training type) from 3 different colonies were observed. We chose newly emerged workers that were active foragers at the time that trials were to begin. Rather than assigning treatments to bees in a random order, we alternated training types in order to distribute them equally among colonies. The behavior of each bee during its last 15 patch visits of training, herein ‘late training’, and all 10 patch visits in the test (as always, excluding the last patch visit of a bout) was recorded to video. For the last 3 bees assigned to each treatment, we also recorded 15 patch visits of ‘early training’, which began after a bee’s initial 10 patch visits (ensuring that they had learned to efficiently handle flowers and travel between patches). One observer used JWatcher v 6.1 (Blumstein and Daniel 2007) to measure the number of rewards
found, inter-reward search times, giving-up times (time from obtaining the last reward or entering the patch, whichever was more recent, to leaving), and travel times among patches.

4.2.2 Setup and parameterization of the optimal foraging model

The model environment corresponds to our experimental training environments but consists of an infinite series of patches. Before a patch is visited, \( y \) of 12 flowers contain a nectar reward, and the distribution of \( y \) among patches is the model analogue of our experimental manipulation. Each patch visited in the HV environment has probability 1/2 of being either a low (\( y = 1 \)) or high (\( y = 9 \)) quality patch. Each patch in the uniform environment initially has five rewards (\( y = 5 \)) with probability one. These are the prior distributions that enter into the optimal foraging model below. Bees are assumed to ‘know’ the appropriate distribution with certainty by late training.

A foraging strategy divides bees’ time into searching for rewards on a patch or traveling among patches. Real bees also had a non-zero handling time associated with nectar consumption, but handling time has no effect on the choice of an optimal strategy (Oaten 1977). Inter-patch travel time is a random variable with mean \( \tau \). Inter-reward search times on patches are assumed to be exponentially distributed random variables with rate parameter \( a(y-x) \), where \( a \) is a constant called the ‘searching efficiency’ (sensu Murdoch and Oaten 1975) and \( x \) is the number of rewards already found in the patch. Under this exponential distribution, \( 1/a(y-x) \) is the mean inter-reward time.

Our foraging model thus requires an estimate of three parameters: the average travel time between patches, \( \tau \), and the average searching efficiency in HV and uniform environments. Average travel time, \( \tau \), was estimated as the average measured time for bees to travel between patches. Searching efficiency parameters were estimated by a survival analysis, which assumed that inter-reward times were exponentially distributed with rate \( a(y-x) \). We graphically checked that inter-reward times were indeed approximately exponentially distributed (Appendix A). For each training environment, \( a \) was estimated by maximum likelihood. Estimation of \( a \) requires that all giving-up times be treated as ‘censored’ inter-reward times (the true time is at least as long, but we cannot know how much longer). We handled these censored observations following Haccou and Meelis (1992, Box 4.16).
4.2.3 Optimal foraging: the potential value model

The optimal foraging model is a slightly modified version of the continuous time, random search model of Green (2006). Following other authors, we call it the ‘potential value model’, which finds a rule that is known to maximize energy intake from any particular resource distribution (e.g., Olsson and Holmgren 2000). Here, we give an intuitive account of the model and reserve mathematical details for Appendix B.

Potential value measures patch quality as the expected number of rewards that a forager will obtain from the remainder of the current patch visit, divided by the expected amount of time spent searching the remainder of the patch. Notice that potential value depends on how the forager decides to leave in the future, a unique problem that requires working backwards through time with dynamic programming (Mangel and Clark 1989). The model assumes that foragers leave a patch when potential value falls below a threshold, $C$. Each choice of $C$ leads to a long-term intake rate,

$$R = \frac{E(G)}{E(S) + \tau},$$

(1)

where $E(G)$ is the total number of rewards that a forager following rule $C$ expects to find in a patch, and $E(S)$ is the total time that such a forager expects to spend searching a patch. It can be shown that the threshold rule that maximizes the reward intake rate in a particular environment, $C_i^*$ ($i = \text{HV or uniform}$), is equal to $R_i^*$, the maximum long-term intake rate for environment $i$ (Green 2006). The optimal thresholds for our parameterized models are given in Figs. 4.1 A and B (dashed lines).

The potential value rule optimally combines current patch information (the number of rewards found and the total time spent searching) with prior knowledge of the resource distribution. In this way, rewards found within a given amount of time can give an optimal forager information about the quality of the patch being exploited and, in some cases, information about the number of rewards remaining. With prior knowledge of the HV environment, an optimal forager is partially informed (uncertain) of patch quality only when finding fewer than two rewards (consistent with being on either a high or low quality patch). If no rewards are found, potential value quickly declines to the leaving threshold (Fig. 4.1 A). Finding one reward, however,
indicates the possibility of being in a high quality patch, and finding a second reward indicates a high quality patch with certainty; in both of these cases, potential value sharply increases upon finding a reward and so should bees’ tendency to stay in the patch (an ‘incremental’ response; cf. Wagge 1979). After finding two rewards in the HV environment (becoming fully informed, and knowing that seven rewards remain) potential value varies only with the number of rewards found (but not with time): each future reward suddenly reduces the potential value (tendency to stay) (a ‘decremental’ response, cf. Driessen et al. 1995). Notice that uniform-trained (U-T) optimal foragers are always fully informed and should behave as if they know that all patches initially contain five rewards (Fig. 4.1 B). As above, each reward found only indicates that the patch is closer to being empty.

4.2.4 Statistical model of behavior: Cox’s proportional hazards

The potential value model is a useful tool for identifying optimal behavior but is less useful for describing the actual behavior of foraging animals. In a given context (number of rewards found and time spent searching), the potential value model predicts that bees with common prior information will behave in exactly the same manner. Hence, optimal behavior is ‘inflexible’ because it is fully determined by optimal use of prior information. However, we expected (and indeed observed) that even a single bee in identical contexts would exhibit behavioral variation. A more flexible statistical model, which makes no assumptions about optimality or the effect of prior information, is required to infer bees' actual tendency to stay on a patch. For this purpose, we used a version of Cox’s proportional hazards model (following e.g., Haccou et al. 1991; Wajnberg et al., 1999). Our strategy was to assess the degree of similarity between the potential value model, which is optimized to experimental reward distributions, and the flexible Cox model, which we inferred from experimental data. If the fitted Cox model were to match the potential value model (as depicted in Fig. 4.1), this would imply that bees used prior information to adaptively adjust their patch-leaving tendencies (i.e., that their behavior is ‘Bayesian-like’).

Our Cox model shares an important feature with the potential value model that allows the inference of Bayesian-like foraging: both models assume that bees' tendency to stay in the current patch is a function of the time spent searching the patch, \( t \), and the number of rewards, \( x(t) \), obtained by time \( t \). Potential value is itself a measure of bees’ tendency to stay in a patch; in the Cox model, the analogous measure is given by the reciprocal of the so-called hazard. In
survival analyses of patch residence time (e.g., Wajnberg et al. 1999), the hazard function is the probability density that a foraging animal will leave its current patch in the next instant. The decision to leave a patch under the Cox model is thus a random event that becomes more or less likely depending on bees' experience in the patch (i.e. on $t$ and $x(t)$). It follows that, unlike the potential value model, the Cox model does not predict a sharp leaving threshold.

The hazard function was modeled as

$$h(t, x(t)) = h_0(t) \exp\left(-\beta x(t)\right),$$

where $\beta_i$ is a coefficient measuring the effect of finding the $i$th reward on the hazard, and $h_0(t)$ is the so-called baseline hazard function, which does not depend on the number of rewards obtained. The number of rewards obtained, $x(t)$, was treated as a time-varying, categorical factor (see Wajnberg 2006). We set $\beta_0 = 0$ to treat zero rewards as the reference category, meaning that the hazard for the zero rewards context is given by the baseline hazard function. Coefficients greater than (less than) zero indicate an increased (decreased) tendency to stay in a patch relative to the zero rewards category. For each of the four datasets, HV training and test and uniform training and test, one categorical-predictor model was fitted.

We used the potential values of Bayesian foragers (Figs 4.1 A, B) to derive qualitative predictions for the reciprocal hazard functions of Bayesian-like foragers (Figs 4.1 C, D). Both models predict an incremental response for HV-T bees that find a first and second reward and a decremental response for U-T bees and for HV-T bees in fully-informed contexts. Figures 4.1 E and 1 F translate the predicted adjustments to the reciprocal hazard function in HV and uniform environments, respectively, to the predicted pattern in the $\beta$ coefficients as rewards are obtained. HV-T bees' coefficients are predicted to increase upon finding the first two rewards (an initial incremental response) and then decrease as more rewards are found (a subsequent decremental response), whereas U-T bees' coefficients only decrease as more rewards are found (decremental response). This pattern of $\beta$ coefficients is required for qualitative agreement between the Cox model and potential value model, and it is our primary test of Bayesian-like foraging.

We also fitted Cox models that treated the number of rewards as a continuous predictor,
where $\beta$ is a single 'slope' coefficient. In this case, one model was fitted to each of the following six datasets: partially-informed HV-T bees (zero to two rewards obtained) in: (1) training and (2) test; fully-informed HV-T bees (two to nine rewards obtained) in: (3) training and (4) test; and U-T bees in: (5) training and (6) test. For the uniform and fully-informed HV models, we expected negative slopes (a consistent decremental response to each reward), whereas for partially informed models, we expected positive slopes (see Figs. 4.1E and 4.1 F).

We used the `coxph` function in the R survival package to fit the ten Cox models. Data were arranged in the counting process formulation of survival times, which allows `coxph` to fit models with time-varying covariates (following Therneau and Grambsch 2000). To account for repeated measures on each bee, we used the `cluster` function, which means that the robust sandwich variance estimator was used, as recommended by Therneau and Grambsch (2000). We checked for proportionality of the hazard functions among reward categories, an important assumption of the Cox model, using the R `cox.zph` function, which implements the testing procedure of Grambsch and Therneau (1994). For all models, we failed to reject the null hypothesis that the hazards were proportional (all $p > 0.05$), indicating that our modeling approach is appropriate. Wald tests were used to assess the overall significance of each model.

### 4.3 Results

#### 4.3.1 Predicted behavior in the training environment

Prediction 1 (Figs. 4.1 E and F). HV-T bees that find a first and second reward should show an incremental response (positive $\beta$ coefficients), but all subsequent rewards should decrease their tendency to stay in the patch. Each reward found by a U-T bee should elicit a decremental response (giving increasingly negative $\beta$ coefficients).

Figures 4.2 A and 4.2 B show the estimated $\beta$ coefficients for HV-T and U-T bees, respectively. These figures resemble fairly closely the predictions in Figs. 4.1 E and 4.1 F, indicating that bees' behavior at the end of training approximated the predictions of our optimal foraging model. One notable exception is that HV-T bees tended to stay longer than predicted after
exhausting patches with nine rewards (the confidence interval for \( \beta_0 \) does not overlap zero; Figure 4.2 A). Because of this exception, we could not detect a significant decremental response of HV-T bees in the fully-informed state using the continuous-predictor model.

Prediction 2. Once in a fully-informed state, the potential value model predicts a sharp leaving threshold based on the number of rewards obtained (a ‘fixed number’ rule). The critical number of rewards is six and three for HV-T and U-T bees, respectively (see the hypothetical patch visits in Figs. 4.1 A and B).

There was a large degree of variation in the number of rewards obtained, indicating that bumble bees certainly did not follow a sharp threshold. However, the mean behaviors—6.2 rewards obtained (95% confidence interval: 5.7, 7.7; \( n = 10 \) bees) by HV-T bees on high quality patches and 3.5 rewards obtained (3.3, 3.7; \( n = 10 \)) by U-T bees—were very close to the optimal values.

Prediction 3. The predicted long-term reward intake rate of an optimal forager (\( R^* \)) in our HV and uniform environments, respectively, was 0.21 and 0.23 rewards/s (Fig. 4.1 A, B; recall that \( C_i = R^*_i \)).

HV-T bees in late training achieved a mean long-term intake rate of 0.16 rewards/s (0.14, 0.18; \( n = 10 \) bees), or 78 % of the rate of an optimal forager. U-T bees achieved a rate of 0.19 rewards/s (0.16, 0.21; \( n = 10 \)), or 82 % of the maximum rate. For comparison, we considered a hypothetical, ‘naive’ strategy of searching each patch for a fixed amount of time, such that the probability of finding every reward would be at least 0.95. The intake rate achieved by this strategy would be only 26% and 35% of the maximum rate in the HV and uniform environments, respectively.

To test whether bees’ foraging success in late training was due to learning, we also measured the intake rate of the individuals observed during early training. The intake rate of these six bees in early training was significantly lower, on average, than their rate at the end of training (mean difference: -0.044 rewards/s (-0.083, -0.0050); paired t-test, \( t_5 = 2.90, p = 0.03 \)). Another remarkable contrast between early and late training was the difference in time spent searching an exhausted (fully emptied) patch before giving up. By the end of training, U-T bees in particular had much shorter giving-up times than U-T bees during early training (Fig. 4.3;
estimated mean in early training: 14.4 s (7.6, 21.1), n = 3 bees; estimated mean in late training: 4.4 s (2.8, 6.0), n = 10 bees).

4.3.2 Patch-leaving decisions in the test environment

Some differences in the behavior of U-T and HV-T bees in the training environments are confounded with the distribution of rewards in those environments. For example, HV-T bees were much more likely than U-T bees to abandon a patch without finding rewards, as predicted by the potential value model. However, this difference in behavior may not reflect a difference in decision-making. It is possible that all bees followed a fixed giving-up time threshold that was simply reached much sooner on low quality patches in the HV environment than it was in the uniform environment (where rewards were relatively easy to find). The purpose of the test environment, then, is to compare patch-leaving behaviors on a ‘common background’. Any differences between HV-T and U-T bees in the test environment must be due to decisions based on alternative prior information, learned during training.

Given the same current information about patch quality in the test environment (number of rewards found), HV-T and U-T bees did indeed show very different tendencies to stay in a patch (Fig. 4.2 C, D). As expected of U-T bees, each reward found tended to decrease their tendency to stay in the patch (a decremental response, as observed in the training environment). In contrast, finding rewards tended to increase HV-T bees’ tendency to remain in a patch. This incremental response of HV-T bees in the test environment can be partly explained by their (now confirmed) tendency to abandon patches before finding a reward. Seven HV-T bees left a test patch without finding rewards (for a total of 13 instances), compared to only one instance by a U-T bee (Likelihood ratio test, weighted by the number of instances, $\chi^2 = 14.0, p < 0.0002$).

Hence, relative to the baseline tendency for HV-T bees to stay in a patch with no rewards, finding rewards consistently increased their tendency to stay.
4.4 Discussion

4.4.1 Testing optimal foraging theory

The first optimal foraging models to include the stochastic element of finding discrete food items in patches (e.g., Oaten 1977; Green 1980; Iwasa 1981; McNamara 1982) revealed that the form of the best patch-leaving rule depends critically on knowing the distribution of food items among patches. More recent techniques (Green 2006) have now allowed us to extend those models to the specific distributions used in our experimental environments, where the theory can be rigorously tested under controlled conditions. By calculating the best rule in both environments, we were able to confirm that Bayesian-like information use is indeed critical to forage successfully (i.e., that a ‘naive’ forager would perform poorly by ignoring information). We were also able to estimate the overall success of bumble bees relative to an optimal Bayesian decision-maker (on average, bees were 80% optimal, according to the potential value model).

Still, even though stochastic optimality models are more realistic than their predecessors (e.g., Charnov 1976), theorists recognize that animal foragers do not actually compute conditional probabilities and cannot realistically follow the deterministic rules of an optimal forager (McNamara et al. 2006). Completing this final link from foraging theory to actual behavior is the first major contribution of the current study, by matching detailed optimal foraging models to a plausible, non-Bayesian behavioral mechanism that mimics the best rule (cf. Pierre and Green 2008). In this way, we have shown a remarkable correspondence between the sharp incremental and decremental responses predicted of an optimal forager and the actual adjustments to bumble bees’ average tendency to remain in a patch, estimated by the statistical model. As expected of bees with knowledge of a uniform environment, finding rewards almost always decreased their tendency to stay in a patch; as for bees with knowledge of a highly variable environment, finding a reward early in a patch visit sharply increased their tendency to stay.

4.4.2 Bayesian-like foraging via learning and memory

The use of a proximate rule that mimics a Bayesian estimator of patch quality does not alone imply the use of learning or memory (e.g., Pierre and Green 2008). Numerous examples of either incremental or decremental responses to resource items have been reported in studies of
parasitoids (reviewed by Wajnberg 2006) and a previous study of bees (Lefebvre 2007). In all of these studies, however, it could be argued that the distribution of resources was sufficiently similar to the foragers’ ancestral environments, where the observed (incremental or decremental) response to food items may have been hardwired by natural selection (see Wajnberg et al. 1999). The experiments reported here are the first to show that individuals of the same species can adapt to very different environments by using both incremental and decremental responses in the appropriate context (highly variable and uniform distributions, respectively).

Hence, the second major advance of the current study is to confirm that bumble bees’ tendencies to remain in a patch are indeed shaped by adaptive learning and memory. The adaptive use of a learned prior distribution (the central feature of Bayesian foraging by learning) was clearly demonstrated in our test environment, where the observed behavioral differences between U-T and HV-T bees must have been due to alternative information learned from past experience. Consider, for example, that HV-T bees that were initially unsuccessful in a test patch were very likely to abandon it. This is an adaptive tactic in their previous HV environment, where finding no rewards early in a patch visit indicates a low quality patch. Abandoning low quality patches in the HV environment is acceptable if the forager knows that high quality patches can be found elsewhere. Uniform-trained bees with the same average experience in a test patch, however, almost always persisted to find rewards. Persistence is adaptive in their previous uniform environment (and in the test environment, incidentally) because if foragers know that every patch is equal, then finding no rewards early in a visit only indicates that five rewards still remain.

The second line of evidence for bees’ adaptive learning is the differences in behavior from early to late training. Bees improved their average foraging success (reward intake rate) in a matter of hours (less than 45 patch visits), and by the end of training, many aspects of their behavior were consistent with our proximate version of the optimal foraging rule. In particular, it is notable that by the end of training, bees in the uniform environment abandoned exhausted patches only about four seconds (on average) after finding the final reward; in the early training, bees’ average giving-up time was more than three times longer. Although the potential value model predicts that U-T bees should never stay on a patch long enough to find all five rewards, the short giving-up times in late training (and the observed decremental response to rewards) does
suggest a degree of certainty that five was the maximum number available. This learned giving-up behavior of U-T bees is striking when compared to a well-known study of Bayesian-like birds (Lima 1984; see also Valone 1992). In that study (over a course of 10 days and hundreds of patch visits), woodpeckers consistently ‘over-sampled’ exhausted patches, usually by searching the entire patch, even though only 6 of 24 units (the equivalent of our flowers) were rewarding.

Admittedly, the bumble bees in our HV environments (both early and late training) similarly over-sampled high quality patches after finding all nine rewards. Accordingly, they did not show a strong decremental response after finding a second reward, implying that, as opposed to optimality assumptions, bees’ prior information by the end of training did not fully match the HV distribution. Because bees encountered high quality patches during only half of their HV training, however, it is possible that their training experience was not sufficient to learn the maximum of nine rewards and the optimal decremental response that U-T bees seemed to learn. Interestingly, a number of authors (e.g., Lima 1984; Valone 1992) have suggested that over-sampling is adaptive in foragers’ typical (ancestral) environments, where the distribution of resources is rarely so predictable as in a controlled experiment. A highly clumped distribution (e.g., the negative binomial), for example, favors an incremental response to every reward found (Iwasa 1981; Olsson and Brown 2006). Our data suggest the intriguing hypothesis that bumble bees’ innate foraging rule, set by evolutionary history, is an incremental response; with enough experience in the appropriate environment, however, they are also able learn a decremental response. An innate incremental response could be adaptive because clumped distributions have been more common in bees’ ancestral environments or perhaps because an incremental response is the best way to learn about a new environment. The current study cannot fully address our hypothesis because bees in early training were fist exposed to extremely rewarding flowers during the floral recognition phase of training. Prior information from evolutionary history was therefore briefly tainted to an unknown extent.

4.4.3 Bayesian-like foraging in an ecological context

Finally, the behavior of nectar foragers has interesting implications in the ecological context of pollination, and our study is relevant in this respect. Some authors have hypothesized that bumble bees’ apparently fixed rules thumb, such as ‘leave a plant after two consecutive empty
flowers’, influence the evolution of floral displays and rewards (e.g., Harder et al. 2004), including the phenomenon of non-rewarding (‘cheating’) flowers (Bailey et al. 2007). Our study demonstrates, however, that bumble bees’ response to non-rewarding flowers is certainly more dynamic than the proposed ‘rule of thumb’. There are at least two situations in which a bee should accept unrewarded search time on a plant: (1) it has experienced a uniform environment in the past and expects to find a certain number of rewards per plant; (2) after experiencing a highly variable environment in the past, the bee finds some rewards on the plant, encouraging it to stay and gain more information about the plant’s true quality.

It becomes a greater challenge to determine the outcome of selection on floral traits when pollinator behavior on a plant with a certain trait value depends on the traits of other plants in the environment. It is now clear that bumble bees in particular have a remarkable ability to learn about the resources in their environment. As a result, bees’ patch-leaving behavior (and the pollination consequences, from a plant’s perspective) should depend on both the mean quality of rewards (Chapter 3) and the distribution of rewards (this study) that they learn from past foraging experience.
4.5 Literature Cited


Green, R. F. 2006. A simpler, more general method of finding the optimal foraging strategy for Bayesian birds. Oikos 112: 274-284.


4.6 Appendices

4.6.1 Appendix A: Graphical assessment of the assumption of exponentially distributed inter-reward times

Our potential value model assumes that inter-reward times are exponentially distributed; here, we give graphical assessments of this assumption (Figs. 4.A.1 and 4.A.2). We expect log-survivorship curves to be straight lines for exponentially distributed inter-reward times. In general, the exponential assumption seems reasonable, especially in the uniform environment. The largest deviation is in the HV environment with nine rewards remaining (i.e. times between entering a patch and finding the first reward).

4.6.2 Appendix B: Mathematical details of the optimal foraging models

4.6.2.1 Initial definitions

The following definitions describe the context of an animal that has been foraging for $t$ seconds in a patch.

Random variables unknown to the forager at time $t$ (parameters):

1. $Y$ is the initial number of rewards in the current patch—or the number of rewards in the patch at time $t = 0$.

2. $I$ is the number of rewards that will be found in the next $\Delta$ seconds (where $\Delta$ is small).

3. $W$ is the inter-reward time (in seconds)—the time the forager spends searching between the last reward and the next reward. However, if at time $t$ the forager has found zero rewards in the patch, $W$ (still referred to as the inter-reward time) is the time between entering the current patch and finding the first reward. Note also that the forager may not remain in the patch long enough after $t$ seconds to find the next reward (censored inter-reward times) or the next reward may not exist at all (the patch is depleted). If the forager does find the next reward, $W$ will become known to the forager in the future when it finds the next reward.
Random variable known to the forager at time \( t \) (data):

4. \( X \) is the number of rewards already found in the current patch.

Deterministic constant known to the forager (hyperparameter):

5. \( a \) is the ‘searching efficiency’ (as defined by Murdoch and Oaten 1975)—see below for more detail.

4.6.2.2 Conventions

Conditional probability density / mass functions are represented in the usual manner. That is, 
\( p(X=x|Y=y) \) denotes the probability (or probability density) that the random variable \( X \) equals \( x \) given that the random variable \( Y \) equals \( y \).

4.6.2.3 Assumptions

The probability density of the inter-reward time, \( W \), is,

\[
p(W = w | X = x, Y = y) = a(y - x) \exp(-a(y - x)w).
\]

That is, \( W \) is exponentially distributed with rate parameter \( a(y - x) \), when \( X = x \) and \( Y = y \).

The prior distributions of rewards among patches in the two environments are,

\[
p(Y = y) = \begin{cases} 
0.5, & y \in \{1, 9\} \\
0.5, & \text{otherwise} 
\end{cases},
\]

(HV)

\[
p(Y = 5) = 1.
\]

(U)

4.6.2.4 Potential value rule

Given the result in the Appendix of Olsson and Holmgren (1998), the likelihood function for the foragers is,
\[ p(X = x \mid Y = y) = \left( \frac{y}{x} \right) \exp(-ayt)(\exp(at)-1)^y. \]  

(A1)

Similarly, we also have,

\[ p(I = i \mid X = x, Y = y) = \left( \frac{y-x}{i} \right) \exp(-ay\Delta)(\exp(a\Delta)-1)^i. \]

By Bayes’ theorem,

\[ p(Y = y \mid X = x) = \frac{p(X = x \mid Y = y)}{p(X = x \mid Y = 1) + p(X = x \mid Y = 9)}. \]  

(HV)

\[ p(Y = 5 \mid X = x) = 1. \]  

(U)

By the law of total probability,

\[ p(I = i \mid X = x) = \sum_{y=0}^{9} p(I = i \mid X = x, Y = y) p(Y = y \mid X = x), \]  

(HV)

\[ p(I = i \mid X = x) = p(I = i \mid X = x, Y = 5) p(Y = 5 \mid X = x). \]  

(U)

These last probabilities can be used in dynamic programming to find the optimal patch-leaving rule (i.e. the potential value rule or the rule that maximizes Eq. 1). In Green’s (2006) notation, they are denoted by \( r(i,t,x) \).

### 4.6.2.5 Similarities with the modeling approach of Green (2006)

The potential value rule modeling approach above is almost a special case of the random-search continuous-time model of Green (2006). In terms of qualitative model behavior, there is no difference between the two approaches. The major difference is how searching is modeled. Our approach was to represent the inter-reward times as exponentially distributed random variables (see Assumptions above). Green’s (2006) approach was to consider waiting times between finding portions of the patch that could potentially contain a reward. He then modeled these waiting times as exponentially distributed random variables. Thus, we directly modeled the
efficiency of foragers at finding rewards, whereas Green (2006) modeled the efficiency of foragers at checking portions of the patch for rewards.

Obviously these two approaches are very similar and the distinction is subtle. To make this distinction clear, consider the types of data that would be required to parameterize Green’s (2006) model versus our model. To parameterize Green’s (2006) model, one would need to know the times that it took for bees to move between flowers. We did not collect this information because of the technical difficulties involved with identifying when a bee actually checked a flower for a reward. In contrast, to parameterize our model, one would only need to know the inter-reward times. These times were easier to measure because the handling time associated with finding a reward made it obvious when a bee had actually obtained a reward.

The only other difference between the two modeling approaches is that Green (2006) measured time such that foragers would cover 95% of the patch in approximately three time units. In contrast, we measured time in units of seconds; this makes it easier to relate our model to experimental data.
4.7 Figures

Figure 4.1
Graphical depictions of the predicted patch-use behavior of high variance-trained (A, C, E) and uniform-trained (B, D, F) bees. Panels A-D give a measure bees' tendency to stay in a patch (y-axis) as a function of time spent searching the patch (x-axis) and the number of rewards found (numbers associated with each grey line) under the potential value model (A, B) and a Cox proportional hazards model (C, D). The solid black lines represent changes in the tendency to stay during a hypothetical patch visit, where the times at which rewards are found (arrows) are the mean times to finding x rewards measured in the late training stage of our experiments. In A and B, strictly optimal foragers leave the patch when their potential value (i.e. tendency to stay) falls below the threshold value, $C_{HV}^*$ or $C_{U}^*$, respectively (dashed lines). In the Cox model, bees' decision to leave is a random event with a likelihood (i.e. hazard function) that depends on $x$ and $t$. Hence, the reciprocal hazard (y-axis; C, D) measures bees' tendency to stay, and it should mimic the tendencies in A and B (as shown) if bees approximate the optimal rule. Panels E and F give the qualitative pattern in estimated Cox model coefficients that we expect to measure if bees do in fact approximate optimal behavior (these predictions correspond to the data in Fig. 4.2). Each coefficient (y-axis) gives a measure of the increased tendency to stay in a patch upon finding x rewards (x-axis).
Figure 4.2

Fitted Cox model coefficients (y-axes) versus number of rewards obtained (x-axis) by high variance-trained (HV-T; A, C) and uniform-trained (U-T) bees (B, C) in the late training (A, B) and test (C, D) environments. Points with error bars (95% confidence intervals) show fits to the models when x is treated as a categorical factor (there is no error at x = 0 because this category was used as the baseline). Three of four categorical-predictor models were significantly different from a model in which all coefficients were zero (HV training model (A), Wald = 1878, p < 0.0001, d.f. = 9; U training model (B), Wald = 76.5, p < 0.0001, d.f. = 5; U test model (D), Wald = 21.3, p < 0.0008, d.f. = 5), and the model for the HV test (C) was very nearly
significant (Wald = 10.8, p = 0.0544). Slopes of the lines are given by the significant $\beta$
coefficients in the continuous-predictor models (they are not least squares-fits to the coefficients
estimated under the categorical predictor models). Three of six models were significant
(partially-informed HV training model (A), Wald = 11.2, $p = 0.0008$, d.f. = 1; U training model
(B), Wald = 28.7, $p < 0.0001$, d.f. = 1; U test model (D), Wald = 7.33, $p = 0.0068$, d.f. = 1). In
general, HV-T bees show an increased tendency to stay upon finding the first rewards in a patch
(A, C), whereas U-T bees show a decreased tendency to stay as each reward is found (B, D).
Figure 4.3

Boxplots of observed giving-up times on exhausted patches (time since finding the last reward to leaving the patch) with either 5 (uniform, U) or 9 (high variance, HV) initial rewards in early and late training. The number of patches that were exhausted and the number of bees that exhausted at least one patch are shown. If, by late training, bees had learned the distribution of rewards in their environments (i.e., had become fully informed), they should ‘know’ that the patch is exhausted and hence search only briefly. This prediction is supported for U-trained bees but not for HV-trained bees.
Figure 4.A.1
Kaplan-Meier estimates of the log-survivorship function (with 95% confidence intervals) for the event of finding the next reward in the HV training environment. Each panel is for a particular number of rewards remaining in the patch; for example, the panel marked '1' gives the log-survivorship curve associated with the time between finding the eighth and ninth rewards. Least-squares lines are plotted for visual reference.
Figure 4.A.2

Kaplan-Meier estimates of the log-survivorship function (with 95% confidence intervals) for the event of finding the next reward in the uniform training environment. Each panel is for a particular number of rewards remaining in the patch; for example, the panel marked '1' gives the log-survivorship curve associated with the time between finding the fourth and fifth rewards. Least-squares lines are plotted for visual reference.
Chapter 5
The origin of gender dimorphism in animal-dispersed plants: disruptive selection in a model of social evolution

5 Abstract

Dioecy (separate sexes) in plants is associated with animal fruit dispersal, but hypotheses for a role of dispersal in the origin of gender dimorphism have received little support. Here, I present a patch-structured model to explore the conditions that favor dimorphism when dispersal is coupled with sex allocation. The model shows that if the proportion of fruits dispersed from a cosexual plant increases with its allocation to fruits (causing accelerating fitness returns from dispersed fruits), disruptive selection can arise when the cost of dispersal is minimal and the correlation among patchmates (i.e., relatedness) is high. In reality, however, the proportion of fruits dispersed from a plant’s patch may decline with further allocation to fruits. Even in this case, novel contexts that lead to disruptive selection on sex allocation are discovered, occurring when dispersal costs are high and relatedness is low, causing accelerating returns from non-dispersed fruits. Hence, surprisingly, gender dimorphism can evolve because female-specialists are better able to escape local competition or better able to succeed in it. Building on the few existing models of disruptive selection on social traits, the mechanisms here show that selection for relaxed local competition (cooperation) can sometimes facilitate diversification and sometimes prevent it.

5.1 Introduction

The adaptive allocation of resources to male and female components of a cosexual organism (sex allocation) and the conditions that maintain cosexuality itself have been established by general theory (Charnov et al. 1976; Charnov 1982; Lloyd 1984). Equilibrium sex allocation

---

strategies occur where the fitness returns from a slight increase in allocation to either gender component are equal (Charnov 1982). The shape of these ‘fitness gain curves’ (fitness with respect to increasing proportional allocation to a particular component) also affects the evolutionary stability of cosexuality in a simple way. If at least one gain curve increases at a sufficiently accelerating rate, then selection becomes disruptive at the equilibrium, favoring the evolution of separate sexes or other forms of gender dimorphism (an example of adaptive evolutionary branching, e.g., de Jong and Geritz 2001). Apart from the genetic details of selfing and inbreeding depression that can favor divergence from cosexuality (e.g., Charlesworth and Charlesworth 1978 a,b), plant ecologists in particular have sought out additional contexts that might cause accelerating fitness gains. More detailed models of sex allocation would be helpful in this respect because the assumptions of the general theory do not describe many realistic ecological contexts.

In particular, most plant populations are spatially structured (see Vekemans and Hardy 2004), where individuals interact in local patches and dispersal among patches is only partial. This population ‘viscosity’ (Hamilton 1964) can result in correlated trait values among patchmates due to shared genealogy (relatedness) and the potential for increased local competition (see West et al. 2002, Gardner and Foster 2008). In a viscous plant population, for example, a variant individual that increases its seed production will tend to compete for local germination sites with related and thus similarly productive patchmates (a form of local resource competition [LRC]; Clark 1978). The potential for intense LRC among relatives appears to cause diminishing female fitness gain curves, thus stabilizing cosexuality (Lloyd and Bawa 1984, Campbell 2000).

Some authors proposed that gender dimorphism can arise if dispersal increases at an accelerating rate with further female allocation, which might be the case if animal seed dispersers prefer plants with relatively large fruit crops (Givnish 1980; Bawa 1980). This idea (herein the ‘dispersal advantage hypothesis’) is especially appealing because it might explain why most dioecious plant species—both angiosperms (Bawa 1980; Renner and Rickles 1995; Vamosi et al. 2003) and gymnosperms (Givnish 1980)—also have fleshy ‘fruits’ (here, a structure used to attract seed dispersers). Yet the hypothesis has been widely dismissed (e.g., Lloyd, 1982; Thomson and Brunet 1990; de Jong and Klinkhamer, 2005) because of evidence that the positive relation between a plant’s fruit number (a metric for female allocation) and the
number of fruits removed by animal foragers is usually not accelerating (reviewed by Carr, 1991; de Jong and Klinkhamer, 2005). Hence, even animal-dispersed plants may be unable to escape local competition, so the correlation between dioecy and frugivory remains puzzling.

The ecology of plant sexual systems is often studied by modeling the population dynamics of a dioecious species that invades an otherwise similar cosexual population (e.g., Heilbuth et al. 2001; Wilson and Harder 2003; Barot and Gignoux 2004). These models are mainly consistent with sex allocation theory: if the dioecious population is to persist, female plants must produce more than twice the number of seeds that cosexuals do, and when dispersal is limited (increasing LRC), the seed production requirement is even greater. A recent variant of this model showed that if increased seed production is coupled with increased dispersal success (a result of frugivore attraction), then the stringent requirement for the invasion of dioecy is relaxed (Vamosi et al. 2007). Furthermore, these authors concluded that even a dispersal advantage with strongly diminishing returns can favor dioecy, which seemed like a surprising contradiction to the logic of sex allocation theory. Vamosi et al.’s model only compares the fitness of static dimorphic and cosexual types, however, not the origin of dimorphism via disruptive selection. It is unclear whether disruptive selection can arise despite diminishing dispersal gains, especially in the social context of a structured population.

Here, I re-evaluate the dispersal advantage hypothesis in a model of social evolution. The model extends the scope of classic patch-structured models in sex allocation theory, which typically treat female dispersal as a fixed parameter (e.g., Hamilton 1967; Taylor 1988). A well-known conclusion from these and similar models of social traits is that the equilibrium strategy can be completely independent of the dispersal rate (e.g., Bulmer 1986; Frank 1986; see also West et al. 2002). This will clearly not apply to the extensions here, where seed dispersal is treated as a correlate of sex allocation. Furthermore, whereas most modeling frameworks in social evolution overlook the evolutionary stability of equilibria (Ajar 2003), here it is necessary to determine whether sex allocation converges to an evolutionarily stable strategy (ESS) or to an evolutionary branching point (the origin of dimorphism). The model confirms that branching always involves accelerating fitness gains. Unexpectedly, however, it does not always involve accelerating returns from dispersal; in these cases, I will show that dimorphism can evolve because female-biased individuals succeed in local competition rather than escape it.
5.2 Modeling Approach

5.2.1 Life history

I consider an infinitely large population that is structured into patches, each supporting $N$ cosexual plants (Wright’s island model of demography, where $N$ is called ‘patch size’ or ‘interaction group size’). Following Hamilton and May (1977), patches are assumed to be uniform, stable in quality, and fully occupied. All plants live for only one season so that generations are non-overlapping. Each individual has the same fixed pool of resources to divide between male and female components of reproduction (pollen and fruits, respectively). Fruit is typically more expensive to produce than pollen, so I assume that the expense of a pollen grain is equivalent to some fraction $\alpha$ of a fruit.

The seasonal cycle begins with the production of a large number of flowers and gametes. A proportion $x$ of $K$ reproductive resources is allocated to the female component (yielding $Kx$ ovules that later become fruits), whereas $1-x$ is allocated to the male component (yielding $K(1-x)$ pollen grains). Before fertilization, a fraction $d_m$ of a plant’s pollen is dispersed widely and randomly to the global population. The remaining fraction of pollen, $1-d_m$, is distributed locally and competes at random for mating success within the natal patch. I assume that every ovule in a patch is fertilized and that each produces a single-seeded fruit (after Vamosi et al. 2007). A proportion $d_f$ of a plant’s fruits are then widely dispersed to the global population, but only a fraction $1-c_f$ find an appropriate patch within the population (hence, $c_f$ is the ‘travel cost’ of fruit dispersal). The full proportion $1-d_f$ of non-dispersed fruits are distributed within the local patch (interaction group). When adult plants die, all seeds in a given patch compete at random for the $N$ germination sites, and the cycle is renewed. This scenario is so far equivalent to the ‘DMD’ (male dispersal, mating, female dispersal) life history analyzed by Wild and Taylor (2004). The key modification here is to treat seed dispersal as a function of sex allocation, $d_f(x)$, to determine how alternative functions affect the equilibrium allocation and the likelihood of evolutionary branching.
5.2.2 Fitness in a cosexual population

Plants are assumed to be diploid, with sex allocation determined by additive gene action. Similar to other evolutionary invasion analyses (see Otto and Day 2006), I consider the fate of a rare variant trait value (and corresponding allele) in an otherwise monomorphic population. Although ‘rare’ with respect to the population as a whole, however, the variant trait may be locally common in a particular patch. Such local assortment usually arises because patchmates are related and thus share the variant allele. The fitness of a focal variant individual (its number of successful offspring in the next generation) will depend on its own sex allocation, $x$, the average allocation in its local patch, $y$ (including itself), and the average population-wide allocation, $\bar{y}$ (the resident strategy). This is a neighbor-modulated view of fitness (or a ‘direct fitness’ approach; Taylor and Frank 1996; Frank 1998), which is particularly suited to analyzing selection on correlated traits within groups (Frank 2007). The approach differs in important respects from an exact genetic model (based on the $R_m$ concept of Metz and Gyllenberg 2001), which counts the number of successful emigrants from a patch that descend over many generations from a single variant (Ajar 2003). The results here will show, however, that a direct fitness approach can provide a useful approximation to the more complex genetic model.

The correlation between an individual’s trait value and its patch’s average value (including $x$) will be measured by $S$, the slope of a regression of $y$ on $x$. The coefficient $S$ thus provides conditional information about $y$, given $x$:

$$y = (1 - S)\bar{y} + Sx$$

(1)

because it measures the expected deviation of $y$ from the resident value ($\bar{y}$) given the focal individual’s deviation (see also Grafen 1985; Day and Taylor 1998). Later, it will be necessary to partition $S$ into components, one due to the focal individual itself and the other due to the correlated traits of patchmates:

$$S = \frac{1}{N} + \left(1 - \frac{1}{N}\right)S_p,$$

(2)
where \( S_p \) measures the correlated change in neighbors’ sex allocation with a change in \( x \). Notice that if \( S_p \) is zero (the focal individual’s trait value is uncorrelated with its neighbors’), \( S \) takes a minimum value of \( 1/N \).

The focal individual has two ‘recipient’ fitness components: surviving offspring produced by its own fertilized ovules (female component; \( W_f \)) and surviving offspring sired on other plants (male component; \( W_m \)). For a cosexual plant, total recipient fitness is

\[
W_f(x, y, y') = v_f W_f + v_m W_m,
\]

where the \( v_i \)’s are the reproductive values of the female and male classes. In models of sex allocation and for diploidy in particular, \( v_m = 1/W_m(y, y, y) \) and \( v_f = 1/W_f(y, y, y) \) (e.g., Pen and Weissing 2000), meaning that a focal plant’s female and male fitness are measured relative to the normal fitness of the respective class.

Female fitness depends on the probability that a focal individual’s fruit obtains a germination site in the local patch and global population. These probabilities (\( p_1 \) and \( p_2 \), respectively) are given by the expected number of empty sites in a patch (\( N \)), divided by the total number of native and immigrant seeds competing in the patch. In the local patch, competing seeds come from non-dispersed seeds of the focal individual, non-dispersed seeds of its \( N-1 \) patchmates (with mean trait value \( y_p = (1 - S_p)\bar{y} + S_p x \)), and dispersed seeds from the global population (given by the first, second and third terms, respectively, in the denominator):

\[
p_1(x, y, y') = \frac{1}{K\alpha} \left\{ \frac{N}{x(1 - d_f(x)) + (N - 1)(y_p)(1 - d_f)(y_p) + N\bar{y}d_f(y)(1 - c_f)} \right\}. \tag{3}
\]

The success of the focal individual’s dispersed seeds, on the other hand, depends only on the resident strategy:

\[
p_2(y) = \frac{1}{K\alpha} \left\{ \frac{N}{N\bar{y}(1 - d_f(y)) + N\bar{y}d_f(y)(1 - c_f)} \right\}
\]

\[
= \frac{1}{K\alpha} \left\{ \frac{1}{\bar{y}(1 - c_f d_f(y))} \right\}. \tag{4}
\]

Altogether, the female fitness component is
\[ W_f(x, y, \bar{y}) = K\alpha x \left\{ (1 - d_f(x))p_1(x, y, \bar{y}) + d_f(x)(1 - c_f)p_2(\bar{y}) \right\}, \quad (5) \]

where the first and second terms describe the success of non-dispersed and dispersed seeds, respectively. Notice that the amount of reproductive resources available for fruit production \((K\alpha)\) cancels out in the final expression of \(W_f\), meaning that fitness depends only on an individual’s proportional allocation to fruits \((x)\).

The focal individual’s male fitness component is the mating success of its pollen (proportional to relative abundance) in the local patch and global population, multiplied by the total number of surviving seeds \(\propto W_f\) that are sired locally or globally (first and second terms, respectively, below):

\[ W_m(x, y, \bar{y}) = \frac{(1 - x)(1 - d_m)}{(1 - y)(1 - d_m) + (1 - \bar{y})d_m} K\alpha y \left\{ (1 - d_f(y))p_3(y, \bar{y}) + d_f(y)(1 - c_f)p_2(\bar{y}) \right\} \]

\[ + \frac{(1 - x)d_m}{(1 - \bar{y})} K\alpha \bar{y} \left\{ (1 - d_f(\bar{y}))p_2(\bar{y}) + d_f(\bar{y})(1 - c_f)p_2(\bar{y}) \right\} \]

where

\[ p_3(y, \bar{y}) = \frac{1}{K\alpha} \left\{ \frac{1}{y(1 - d_f(y)) + \bar{y}d_f(\bar{y})(1 - c_f)} \right\}. \]

Note that in the local patch (first term of eq. 6), the focal individual’s mates have an average trait value of \(y\) (producing seeds with a probability \(p_3\) of acquiring a local site), whereas in the global population (second term of eq. 6), all mates have the resident value \(\bar{y}\).

5.2.3 The fitness gradient

The direction of adaptive evolution can be described as the change in total recipient fitness with respect to small variants in the genetically transmitted value (breeding value, \(g\)) of the focal individual’s sex allocation. Following Frank (1998), the marginal change in \(W_i\) is
\[
\frac{dW_t(x,y,\bar{y})}{dt} = v_f \left( \frac{\partial W_f}{\partial x} + \frac{\partial W_f}{\partial y} S \right) + v_m \left( \frac{\partial W_m}{\partial x} + \frac{\partial W_m}{\partial y} S \right)
\]  
(7)

(see the Appendix for a full derivation and cf. eq. 17 in Ajar’s [2003] exact genetic model).

Equation (7) shows that the marginal change to each component of fitness can be partitioned into a direct effect of the change in the focal individual’s own sex allocation \( \frac{\partial W_f}{\partial x} \) and a correlated (‘indirect’) effect of the average allocation in its group \( \frac{\partial W_f}{\partial y} \), where \( S \), the correlation between individual and group, acts to translate the indirect effects into units of fitness for the focal individual.

The approach so far has been general in the sense of excluding the genetic and demographic details of kinship. Yet it can be useful to define the correlation among patchmates as \( r \), a relatedness coefficient from inclusive fitness theory, which emerges from the demographics of the model. In the current model, where the ‘mother’ plant controls the sex allocation strategy, \( r \) will be the coefficient of consanguinity (CC; the probability that random alleles are identical by descent) between a mother and a random patch offspring born on ‘her’ patch, relative to the CC between a mother and her own offspring. Assuming an equilibrium population and weak selection, these coefficients can be calculated from standard recursion relations (e.g., Taylor 1988). For a population of diploid individuals with a DMD life history,

\[
r = \frac{4}{4N - k_f^2 (1 + k_m)^2 (N - 1)}
\]  
(8)

(Wild and Taylor 2004; details in the Appendix), where the \( k_i \)’s measure the probability that a random fruit \( (i = f) \) or pollen grain \( (i = m) \) competes in its local patch. Here, they are

\[
k_f = \frac{1 - d_f(\bar{y})}{1 - d_f(\bar{y}) + (1 - c_f) d_f(\bar{y})} = \frac{1 - d_f(\bar{y})}{1 - c_f d_f(\bar{y})} \quad \text{and} \quad k_m = 1 - d_m.
\]  
(9 a, b)

The correlation among patchmates measured by \( r \) is thus an explicit function of patch size and dispersal parameters, whereas in the general approach, the association between dispersal and \( S \) is left unspecified.
5.2.4 Stability of equilibria

An equilibrium sex allocation strategy occurs where \( \frac{dW_t}{dg} = 0 \), evaluated at the population mean \( (x = y = \bar{y}) \). The stability of an equilibrium depends on two independent properties: convergence stability, ensuring that selection pushes the trait towards the equilibrium when displaced in either direction (Eshel 1983), and invasibility. A convergence stable equilibrium that is non-invasive (implying a fitness maximum) is an ESS (herein \( y^* \)), whereas a convergence stable but invasive equilibrium (implying a fitness minimum) is a branching point.

Because analytical solutions for equilibria can not be obtained when dispersal and sex allocation are correlated, I evaluated the convergence stability and invasibility of equilibria with pairwise invasibility plots (PIPs), a common numerical tool in evolutionary invasion analyses (e.g., Otto and Day 2006). The analyses require that \( W_t(x, y, \bar{y}) \) be expressed only in terms of the rare variant and the resident trait value, but I have effectively done so by defining \( y \) (or \( y_p \)) as a function of \( x \) and \( \bar{y} \) (eq. 1). By using \( S \) or the neutral relatedness defined by eq. 8, I also assume weak selection on sex allocation and no effect of selection on \( S \) or \( r \).

It is important to note that the invasibility condition from a direct fitness approach is known to differ from the condition from an explicit genetic model. Under the assumptions of weak selection and neutral relatedness, Day and Taylor (1998) show that differences between direct fitness and the fitness function in a genetic model can lead to a slight difference in the invasibility condition. Furthermore, an exact evaluation of invasibility requires an evaluation of the effect of selection on relatedness. Ajar (2003) presents an exact genetic model that incorporates such effects, but numerical evaluation of its fitness function \( (R_m) \) can become computationally unwieldy, especially for large patch sizes. In Appendix B, I show representative examples from the current model that illustrate a close correspondence between \( W_t \) and \( R_m \), especially when considering only local invasibility at an equilibrium. This close correspondence justifies the use of direct fitness in the results below.
5.3 Results

5.3.1 Prelude: fruit dispersal fixed

I begin by treating fruit dispersal as a fixed parameter to confirm that the model agrees with existing analytical results. To simplify matters, I assume no cost of fruit dispersal and consider two extreme scenarios: only local distribution of pollen ($d_m = 0$, with no inbreeding depression) or complete dispersal of pollen ($d_m = 1$). In the first scenario, setting the fitness gradient (Eq. 7) to zero and defining $S$ as $r$ (Eq. 8) gives the equilibrium

$$y^* = \frac{1 + N}{2N},$$

which corresponds to Hamilton’s (1967) famous result for local mate competition (LMC) and is a case in which $y^*$ is completely independent of $d_f$. In this case of complete LMC, the equilibrium is increasingly female-biased (less locally-competitive) as patch size decreases because smaller $N$ acts to increase relatedness and thus the potential for intense local competition among pollen grains. A decline in $d_f$ might be expected to counteract selection for female bias because this extra viscosity leads to more local competition among seeds. However, the decline in $d_f$ also increases $r$, and these two effects cancel exactly. Gardner et al. (2009) present a detailed analysis of this model and some extensions.

The second scenario, with full dispersal of pollen and local competition among fruits, is less well-studied and of greater interest for our purposes. Beginning with the general approach, the equilibrium sex allocation is

$$y^* = 1 - \frac{1}{2 - (1 - d_f)^2 S},$$

which shows that, if all else is held constant, the equilibrium is increasingly male-biased (less locally competitive) with greater correlation among patchmates but becomes more female, approaching $y^* = 0.5$, as fruit dispersal increases (alleviating LRC; Fig. 5.1 A). It is again useful to specify the correlation among patchmates as $r$, and after making this replacement in eq. 11,
\[ y^* = \frac{1}{2} \left( \frac{3 - 3N + df(d_f - 2)(N + 3)}{1 - 3N + df(d_f - 2)(N + 1)} \right). \] (12)

While this expression is more cumbersome than eq. 11, Fig. 5.1 B shows that allocation to fruits increases towards \( y^* = 0.5 \) as patch size increases and at higher values of fruit dispersal (both acting to reduce relatedness [Fig. 5.1 C] and hence the potential for intense LRC). These results confirm that selection biases sex allocation toward the gender component that is least likely to face increased local competition on the natal patch (e.g., Wild and Taylor 2004) and that the bias diminishes as the extent of LRC and LMC become similar (here, as \( df \) and \( dm \) converge to one).

A large number of numerical examples suggest that when \( df \) is a fixed parameter, all equilibria are both convergence stable and non-invasible, thus maximizing individuals’ (neighbor-modulated) fitness (although the equilibrium is neutral when \( dm = df = 1 \)). One reason for non-invasibility is that when \( df \) is constant, the focal plant’s absolute number of dispersed fruits (given by \( df Kx \)) is always linear in \( x \) (i.e., never accelerating). Stabilizing selection at \( y^* \) is particularly strong with partial pollen dispersal \( (dm < 1) \) because LMC causes decelerating male fitness returns. Hence, to characterize the conditions that can lead to disruptive selection, the following analyses of selection on \( x \) will continue the \( dm = 1 \) scenario.

5.3.2 Case 1: dispersed proportion of fruits increases with fruit production

This case is motivated by Vamosi et al.’s (2007) assumption that the proportion of fruit dispersed from a plant increases with fruit production, though with diminishing returns. In their model, the critical function is the increase in dispersal experienced by unisexual females that produce \( \beta \) times more seeds than a cosexual type. This is not directly comparable to the gain curves in classical sex allocation theory (which range from 0 to 100% allocation), and indeed, this is a key difference between the two approaches. Nevertheless, it is informative to explore a naive translation of Vamosi et al.’s assumption in a model of sex allocation. An important first result is that whenever \( df(x) \) is strictly increasing, the focal plant’s absolute number of fruits dispersed increases with \( x \) at an accelerating rate (Fig 5.2 A, B).
Hence, even if \( d_f(x) \) increases with diminishing returns, there are indeed conditions that can drive sex allocation to an evolutionary branching point. I first consider two examples that use a power function, \( d_f(x) = x^a \) with \( a < 1 \) (Fig. 5.2 A, i and ii), which is a standard method of modeling gain curves in sex allocation theory. For \( a = 0.2 \) and 0.8, respectively, Figs. 5.3 A and B indicate where the equilibrium sex allocation is an ESS or a branching point at varying combinations of \( c_f \) and \( N \). In general, branching points arise when the cost of fruit dispersal is minimal and the size of interaction groups is small (implying high relatedness, or correlation among patchmates). For small patches in particular \((N < 5)\), branching points arise across a wider range of dispersal costs when dispersal gains accelerate at a relatively fast rate (compare Fig. 5.3 B to 5.3 A).

The conditions that cause disruptive selection in Case 1 can be understood by examining the components of the female fitness gain curves due to success in local and global competition (Fig. 5.4). Even though fitness gained through dispersal increases with female allocation at an accelerating rate (solid lines), Fig. 5.4 shows that if the cost of fruit dispersal is high (red curves), total female fitness at an equilibrium is mainly influenced by local success (dashed lines) (note that total female fitness at equilibrium equals 1). Only when \( c_f \) is low (blue curves) do the accelerating gains from dispersed seeds have a large enough influence on the total female gain curve to introduce disruptive selection at an equilibrium. Furthermore, when the correlation among patchmates is high (Fig. 5.4 A), the local fitness gain curves are relatively flat (compared to the hump shaped curves in Fig. 5.4 B) because a variant that increases \( x \) gains little by competing with itself or with similarly productive patchmates. When added to the accelerating returns from dispersed seeds, the total female gain curve is more likely to be accelerating when \( S \) is high than when low.

In the final example, I consider a function giving accelerating numbers of dispersed fruit but a relatively low average dispersal rate (Fig. 5.2 A, iii), which selects for sex allocation equilibria that are always locally evolutionarily stable (Fig. 5.3 C). This result arises because when dispersal is incomplete (meaning fewer immigrant seeds in the natal patch), further allocation to fruits yields increasing but decelerating local fitness gains from non-dispersed seeds. As above, this acts to counter the accelerating returns from dispersal.
Overall, disruptive selection in Case 1 tends to arise when sex allocation is first driven to 50% female allocation (or lower, if patch size is small; Fig. 5.3 A-C). At these points, variants that are slightly more male or slightly more female can both invade, where the advantage for the more-female variants is to escape intense LRC.

5.3.3 Case 2: dispersed proportion of fruits declines with fruit production

In contrast to Case 1, whenever \( d_f(x) \) is strictly decreasing, the focal plant’s absolute number of fruits dispersed increases with diminishing returns (Fig 5.2 C, D). The simplest way to model a negative function for \( d_f(x) \) is with a linear decline (Fig. 5.2 C, iv and v), where the intercept of the line represents the ‘baseline’ dispersal rate, \( b \). For examples with \( b = 0.9 \) and 0.5, respectively, Figs. 5.3 D and E indicate the stability properties of the sex allocation equilibria at varying combinations of \( c_f \) and \( N \). Here, the main result is that branching points are more likely to arise when the baseline dispersal rate is high and patch size is large (both tending to decrease the correlation among patchmates, or relatedness) and, in contrast to Case 1, when the cost of fruit dispersal is intermediate to high.

Again, the local and global components of the female fitness gain curves (Fig. 5.5) help to explain why disruptive selection arises in Case 2. As above, when the cost of dispersal is high (red curves), the total female gain curve is strongly influenced by fitness gained in local competition (dashed lines). Now, however, the local fitness gain curve can increase with accelerating returns if the correlation among patchmates is small (Fig. 5.5 B). This is understood by noting that when \( c_f \) is large, there are few immigrants entering any given patch, decreasing the total number of seeds competing for local sites. A focal individual is more likely to succeed in local competition when the correlation among patchmates is small, meaning that a variant individual that increases its allocation to fruits will tend to compete against patchmates with the less competitive resident trait. On the other hand, if \( S \) is high, the local component of the gain curves are either decelerating or only weakly accelerating because of intense LRC among similarly competitive patchmates (dashed lines in Fig. 5.5 A).

The final example in Case 2 (Fig. 5.2 C, vi) gives relatively low average dispersal rates but results in evolutionary branching in a parameter space that is comparable to the example with \( b = 0.9 \) (compare Fig. 5.3 F to 3 D). The simple explanation is that when the number of fruits
dispersed is strongly decelerating in $x$ (Fig. 5.2 D, vi), the number of fruits staying in the local patch is strongly accelerating, and it is the local fitness gain curve in Case 2 that causes disruptive selection. All examples in Case 2 have been limited to strictly decreasing functions for $d_f(x)$, but it should be noted that hump-shaped functions for $d_f(x)$, which also translate into mostly diminishing gains in absolute fruit dispersal, can lead to even stronger accelerating fitness gains than are evident in Fig. 5.5.

Overall, disruptive selection in Case 2 tends to arise when sex allocation is first driven near to and beyond 50% female allocation (Fig. 5.3 D-F). At these points, variants that are slightly more male or slightly more female can both invade, where the advantage for the more-female variants is to succeed in competition for local germination sites.

5.4 Discussion

The above analyses establish two very different causes of disruptive selection on sex allocation, and for both cases, understanding the social context of selection is critical. The first potential route to gender dimorphism (in Case 1) corresponds to the original dispersal advantage hypothesis (Givnish 1980; Bawa 1980), which predicts the adaptive evolution of dimorphism when female-specialists are better able to escape competition for local resources. Here, I add that for disruptive selection to arise under this hypothesis, the cost of fruit dispersal must be small (so that extra dispersal gains translate to fitness gains) and that disruptive selection should be strongest when competition is localized to small interaction groups. The second route to dimorphism (in Case 2) is a new finding, which surprisingly depends on the ability of locally-distributed fruits to re-colonize their natal patch (herein the ‘local advantage hypothesis’). In this case, disruptive selection on sex allocation arises when the cost of dispersal is moderately high and when the correlation among patchmates (relatedness) is low (primarily a result of large patch size). Interestingly, the first route to dimorphism (via dispersal advantage) requires the existence of localized competition and population viscosity from which female-specialists can escape. These same two factors, however, actually inhibit the second route to diversification (via local advantage) because they intensify local competition for resources whenever a more-female variant arises.
5.4.1 Plausibility of the hypotheses

Theory for the role of dispersal in the evolution of gender dimorphism, as with any past evolutionary event, can be difficult to test (Charlesworth, 1996). Furthermore, where animal-dispersed fruits are correlated with dioecy in plants, it is not always clear that the mode of dispersal preceded the origin of dimorphism (Vamosi et al. 2003). Nevertheless, the value of the theory is to clarify all possible outcomes and the conditions in which those outcomes might arise in nature.

The dispersal advantage hypothesis is theoretically plausible if the proportion of fruits dispersed from a plant’s patch increases with its allocation to fruits. This situation seems to arise only rarely, however, based on evidence of the relationship between fruit removal and fruit crop size in nature (see Laska and Stiles 1994; de Jong and Klinkhamer 2005). Even in cases where this relationship is positive, data on fruit removal gives only limited information about the ultimate distribution of dispersed seeds (the realized ‘seed shadow’; e.g., Dennis and Westcott 2007). Especially in dense patches, frugivores tend to move only short distances, thus distributing seeds to neighboring sites within the patch (Carlo and Morales 2007). Furthermore, large fruit crops can attract a diversity of frugivores, but many of those visitors simply drop the fruits in the vicinity of the parent plant (‘seed waste’; Howe 1980). These observations point to the plausibility of the local advantage hypothesis, which requires that the proportion of dispersed fruits decreases with further allocation to fruit production. This condition is also consistent with a number of studies that report a negative relation between crop size and the proportion of fruit removed (e.g., Murray 1987 and refs. therein), as long as non-removed fruits are eventually distributed within the local patch.

If patch sizes are not large (e.g., \( N < 10 \)), evolutionary branching under the local advantage hypothesis can require the combination of costly dispersal and a high (inter-patch) dispersal rate at a potential branching point. Yet it seems possible that when dispersal is costly, any additional trait that reduces the level of dispersal should be favored. A trait such as smaller fruit size, for example, might reduce long-distance dispersal (see Seidler and Plotkin 2006), thus affecting the level of local competition and, potentially, the size of the local interaction group. It may therefore be interesting to consider the joint evolution of sex allocation and a trait such as fruit
size to determine whether additional selection on the dispersal rate can eliminate branching via local advantage.

5.4.2 Limitations and additional extensions

The models here are simplified in a number of ways, sometimes limiting their scope to particular scenarios. First, by assuming weak selection on sex allocation, branching points were defined as equilibria where small variants on either side of the equilibrium have higher fitness than the resident. Hence, if some genetic mechanism allows two discrete gender morphs to coexist (i.e., by minimizing recombination between them), then a population at a branching point may diverge by gradual steps. This is consistent with a gradual transition from monoecy (cosexual plants with unisexual flowers) to dioecy by way of slight, divergent adjustments in the number of male and female flowers on a plant, which is thought to be a common evolutionary pathway (e.g., Lloyd 1975; Renner and Ricklefs 1995; Freeman et al. 1997). Dioecy can also evolve by mutations of large effect (see Barrett 2002), and disruptive selection at a branching point is likely to be involved in this case as well. A complete analysis of this pathway, however, would require further numerical evaluations of an exact genetic model.

Inbreeding depression is a key factor that was omitted from the models. The issue of inbreeding was avoided by assuming complete pollen dispersal from the patch (and thus linear male gain curves), which may sometimes approximate the conditions of abiotic pollination (reviewed by de Jong and Klinkhamer 2005) or cases in which cosexual plants already have mechanisms to avoid selfing. Inbreeding has two important effects, however, that might affect the outcome of the above analyses. First, selection to avoid severe inbreeding depression is expected to impose disruptive selection on sex allocation (Charlesworth and Charlesworth 1978 a,b), so it is generally understood that ecological factors (e.g., LRC) and inbreeding avoidance have both contributed to the evolution of dimorphism (e.g., Freeman et al. 1997; Charlesworth 1999). However, the second effect of inbreeding is to elevate relatedness within patches, and for Case 2 in particular, this is expected to stabilize cosexuality. If disruptive selection on sex allocation is to arise by the local advantage hypothesis, it seems that the disruptive effects of inbreeding avoidance would have to outweigh the stabilizing effects of higher relatedness.

Another possible extension to consider is the consequence of social interactions among plants beyond the competition for germination sites. When attracting frugivores, for example, the
attractiveness of a focal individual’s allocation to fruits may depend on the average allocation of patchmates. Indeed, there are examples in the literature of both competitive and facilitative interactions at the frugivore attraction stage (reviewed by Carlo et al. 2007), which could be incorporated into the focal individual’s dispersal function. Ultimately, the origin of dimorphism will depend on the shape of the relationship between allocation to fruits and dispersal gains, but additional interactions could select for higher or lower average dispersal rates, which, as shown above, can also influence the likelihood of evolutionary branching.

5.4.3 Comparison with Vamosi et al. (2007)

As with all previous models of sex allocation, the current study confirms that the gradual evolution of gender dimorphism is only possible when individuals gain accelerating fitness from further allocation to a particular gender function (e.g., Charnov et al. 1976). In Case 1, accelerating gains from dispersed seeds inevitably arise and can often lead to disruptive selection. Even in Case 2, where I imposed diminishing gains from dispersed seeds, disruptive selection for dimorphism arises because of accelerating fitness gains from locally distributed seeds. Vamosi et al. (2007) found that diminishing dispersal gains can facilitate the coexistence of dioecious and cosexual strategies, but it is important to recognize that this result involves neither of the mechanisms explored in the current study.

Vamosi et al.’s (2007) approach can be interpreted as introducing a purely dimorphic strategy (two mutations of large effect) into a cosexual population. Their model shows that a (diminishing) dispersal advantage associated with increasing seed production in pure females (increasing by $\beta$ times the production of coesexuals) can give the dimorphic strategy twice the fitness of cosexuals, even when $\beta < 2$. Yet when considering large mutational steps, the fact that the dispersal advantage is diminishing does not seem to matter as much as the magnitude of the advantage does. When dimorphism evolves by gradual steps, on the other hand, the shape of fitness returns is critical, and the current paper has established the possible outcomes.

5.4.4 Evolutionary branching in social evolution

There are surprisingly few examples of disruptive selection in models with spatial and genetic structure and even fewer in which the roles of population viscosity and relatedness are clear. Of those that exist, however, there is an emerging consensus that limited dispersal and localized
interactions (leading to high relatedness) can inhibit branching (reviewed by Lion and van Baalen 2008). This seems to occur in situations where divergence involves increased investment in local competition among patchmates; in that case, high relatedness among patchmates selects for relaxed investment in local competition (i.e., cooperation), which in turn inhibits diversification (see Day 2001).

Results from the current study, however, illustrate that selection for cooperation can either inhibit or facilitate diversification, depending on the ecological context. Case 2 is similar to existing examples (e.g., Day 2001; see also Ajar 2003) because, in that context, allocation to female function leads to an increase in competition for local resources. High relatedness selects for cooperation in the form of increasingly male-biased allocation (recall that the male component, pollen, never competes locally), which in turn inhibits diversification. In the context of Case 1, however, allocation to female function leads to an escape from local competition. Hence, in this case, allocation to female function can be a cooperative act, and as expected, high relatedness can facilitate diversification.

Additional examples of disruptive selection on social traits may be scarce because the evaluation of evolutionary stability in models of kin selection (including the neighbor-modulated [direct] fitness approach) has been unclear (see Doebeli 1999; Doebeli and Hauert 2006). In fact it is known that the evolutionary stability conditions from a direct fitness approach generally fail to match a simplified genetic model (Day and Taylor 1998), but the extent of the failures have not been emphasized. Ajar (2003) has since used the $R_m$ framework to derive an exact genetic model for the analysis of disruptive selection on social traits. Using this framework as the benchmark, the current study demonstrates that a direct fitness approach can closely approximate fitness (and the evolutionary stability conditions) in the exact model. This is important because a direct fitness formulation is often simpler than the exact genetic model, especially when numerical computations are required. Furthermore, as a heuristic approach, kin selection models often make the biological details of social adaptations more transparent (Day and Taylor 1998). The current paper provides one example, illuminating the complex ecological details that can underlie disruptive selection on plant sex allocation.
5.5 Literature Cited


5.6 Appendices

5.6.1 Appendix A: Derivation of the fitness gradient in a monomorphic population

Here, I derive the marginal change in fitness with respect to an individual’s transmitted breeding value \( g \) (Equation 7 in the main text). Because the average phenotype in the group \((y)\) is a correlated trait of the focal individual \((x)\), both \(x\) and \(y\) are treated as functions of \(g\). Using the chain rule, the expansion of \(dW_t/dg\) is

\[
\frac{dW_t(x, y, \bar{y})}{dg} = v_f \left( \frac{\partial W_f}{\partial x} \frac{dx}{dg} + \frac{\partial W_f}{\partial y} \frac{dy}{dg} \right) + v_m \left( \frac{\partial W_m}{\partial x} \frac{dx}{dg} + \frac{\partial W_m}{\partial y} \frac{dy}{dg} \right),
\]

where the partial derivatives are evaluated at \(\bar{y}\), and in the second step, each phenotypic derivative is interpreted as the slope of a regression of actor phenotype on recipient genotype (Taylor and Frank, 1996; Frank 1998). The first coefficient, \(\tilde{r}\), measures the slope of mother phenotype on offspring genotype; the second, \(\tilde{R}\), measures the slope of average group phenotype on offspring genotype. Following Frank (1998, p. 75), I assume that \(R = \tilde{r} S\), where \(S\) is the phenotypic correlation between the focal individual’s phenotype and the average phenotype in the group. Making this substitution in (A1) and dividing the entire RHS by \(\tilde{r}\) gives eq. 7 in the main text.

If the correlation among patchmates is due to kinship alone, then the regression coefficients in eq. A1 can be interpreted differently. In this case, I replace \(\tilde{r}\) with \(R\), the coefficient of consanguinity (CC) between a focal plant and its own offspring, and \(\tilde{R}\) with \(\bar{R}\), the CC between the focal plant and a random offspring born on its patch (following the notation of Wild and Taylor 2004). Replacing the regression coefficients in (A1) with the appropriate CCs and dividing the RHS by \(R\) leaves

\[
\frac{dW_t(x, y, \bar{y})}{dg} = v_f \left( \frac{\partial W_f}{\partial x} + \frac{\partial W_f}{\partial y} \frac{\bar{R}}{R} \right) + v_m \left( \frac{\partial W_m}{\partial x} + \frac{\partial W_m}{\partial y} \frac{\bar{R}}{R} \right).
\]
The ratio $\bar{R}/R$ is a type of within-patch relatedness coefficient that has been used by many authors (e.g., Frank 1986; Taylor 1988). Wild and Taylor (2004) calculated its component CCs, assuming diploidy and a DMD life cycle:

\[
\bar{R} = \frac{1}{2} \left[ \frac{2}{4N - 2k_m - k_j^2(1 + k_m)^2(N - 1)} \right] 
\]

\[
R = \frac{1}{2} \left[ \frac{4N - k_j^2(1 + k_m)^2(N - 1)}{4N - 2k_m - k_j^2(1 + k_m)^2(N - 1)} \right].
\]

(A3)

In the main text, $\bar{R}/R = r$ is given as Equation 8.

5.6.2 Appendix B: Comparing the neighbor-modulated fitness approach with an exact genetic model

Here, I present the sex allocation model from the main text in the $R_m$ framework of Metz and Gyllenberg (2001), closely following the description of an exact computation of $R_m$ from Ajar (2003). As in the main text, the model is based on the infinite island population structure and assumes diploidy and additive gene action. Although Ajar (2003) considers a haploid population, my assumption of complete pollen dispersal (used here and throughout Cases 1 and 2 in the main text) restricts attention to heterozygote variants, making the two approaches effectively equivalent.

$R_m$ is the overall production of successful emigrants from a patch, descended over several generations from a single variant individual (Ajar 2003). An exact computation of $R_m$ follows the distribution of the number of these descendants. Each generation, a large pool of seeds compete at random for the $N$ germination sites in a patch. The probability that a randomly chosen seed descends from one of the $j$ variants (with sex allocation $x$, in a population with allocation $\overline{y}$) in the focal patch is

\[
\pi_j(x, \overline{y}) = \frac{1}{2} \frac{jx(1 - d_j(x))}{jx(1 - d_j(x)) + (N - j)\overline{y}(1 - d_j(\overline{y})) + N\overline{y}d_j(\overline{y})(1 - c_j)}
\]

(1)
(cf. Eq. 3 in the main text), where the 1/2 represents the variant’s genetic contribution to seeds. The probabilities, \( a_{qj} \), that there are \( q \) variant descendants from \( j \) variant plants in the patch are

\[
a_{qj} = C_N^q (\pi_j)^q (1 - \pi_j)^{N-q},
\]

where \( C_N^q \) is the binomial coefficient \( N!/ [q!(N-q)!] \). Let \( A = (a_{qj}) \) for \( q = 1, ..., N, j = 1, ..., N \) so that the probability of \( j \) variants being in the patch at generation \( t \) is \( p_j = (A^t i)_j \), where \( i \) is the transpose of \((1, 0, ..., 0)\), a vector that represents the initial distribution of the variant.

In each generation, the expected number of successful emigrant offspring from each variant plant is

\[
h_j(x, \gamma) = \frac{1}{2} \left( j x d_j(x)(1 - c_j) \frac{1}{\gamma(1 - d_j(\gamma)) + yd_j(\gamma)(1 - c_j)} \right) + \frac{1}{2} j \left( 1 - \frac{x}{1 - y} \right),
\]

where the first term represents successful dispersing seeds (cf. Eq. 4 in the main text), and the second term represents the siring success of dispersed pollen. Letting \( h \equiv (h_j) \), the expected outflow of successful variants from the patch is

\[
\sum_{j=1}^{N} h_j p_j = h A^t i.
\]

By summing over generations, the long term growth rate of the variant trait (allele) is

\[
R_m(x, \gamma) = h \sum_{t=0}^{\infty} A^t i = h (I - A)^{-1} i.
\]

Numerical calculations of \( R_m \) were computed in pairwise invasibility plots for select parameters in Case 1 and Case 2 of the sex allocation model. The results, and a comparison to the neighbor-modulated fitness approach, are given in Figs. 5.A.1 and 5.A.2. In general, the figures illustrate a close correspondence between the two methods. Most importantly, they show the same shift from local evolutionary stability to local instability that is illustrated in Fig. 5.3 in the main text.
5.7 Figures

Figure 5.1
(A) The evolutionarily stable (ES) sex allocation strategy (proportional allocation to female function) as a function of the correlation among patchmates at varying degrees of fruit dispersal ($d_f$, the proportion of fruits dispersed from an individual). In (B), the correlation among patchmates is defined explicitly as a relatedness coefficient (equation 8 in the main text) so that $y^*$ can be plotted as a function of interaction group (patch) size, $N$, at varying levels of $d_f$. (C) The change in the relatedness coefficient ($r$), which measures genetic similarity among patchmates, with increasing $N$. For all panels, $d_m = 1$, $c_f = 0$. 
Figure 5.2

The proportion of fruits dispersed from an individual (A, C) and the corresponding absolute number of fruits dispersed (B, D) as a function of an individual’s proportional allocation to fruits \(x\). Six examples for \(d_f(x)\) are considered: (A) \(i = x^{0.2}, ii = x^{0.8}, iii = 0.5x^{0.5}\); (B) \(iv = 0.9-0.2x, v = 0.5-0.2x, vi = 0.7-0.5x^2\). In these examples, plants are assumed to have a resource pool capable of producing 100 fruits \((K = 100, \alpha = 1)\).
Figure 5.3
Classification of sex allocation equilibria as local evolutionarily stable strategies (ESSs) or local evolutionary branching points (shaded) at varying combinations of interaction group (patch) size and dispersal cost. The numbers within individual cells give the sex allocation equilibria as percent female allocation. Evolutionary stability of equilibria was determined by investigating pairwise invasibility plots (examples given in Appendix B).
Figure 5.4

Female fitness gain curves at equilibrium, partitioned into components due to fitness gained locally (via non-dispersed seeds) or globally (via dispersal), in cases where the correlation among patchmates is high (A) or low (B). Results are from Case 1, with $d_f = x^{0.8}$ (corresponding to Fig. 5.3 B). Sex allocation equilibria (blue arrows for $c_f = 0$ and red arrows for $c_f = 0.8$) occur where the local and global fitness gains add to 1.
Figure 5.5

Female fitness gain curves at equilibrium, partitioned into components due to fitness gained locally (via non-dispersed seeds) or globally (via dispersal), in cases where the correlation among patchmates is high (A) or low (B). Results are from Case 2, with \( d_f = 0.9-0.2x \) (corresponding to Fig. 5.3 D). Sex allocation equilibria (blue arrows for \( c_f = 0 \) and red arrows for \( c_f = 0.8 \)) occur where the local and global fitness gains add to 1.
Figure 5.A.1

Pairwise invasibility plots for selected parameter combinations of a model in Case 1 (with $d(x) = x^{0.8}$), solved using the neighbor-modulated fitness approach (a-c) and an exact genetic model (d-f). Grey regions indicate that the variant trait value (vertical axis) has higher fitness than the resident (horizontal axis). Lines separating grey and white regions indicate where variant and resident fitnesses are equal, and the equilibrium of the model is where these lines cross. All cases of branching point equilibria are indicated with arrows. In general, these examples show that neighbor-modulated fitness is a close approximation to fitness in the exact genetic model.
Figure 5.A.2

Pairwise invasibility plots for selected parameter combinations of a model in Case 2 (with $d_f(x) = 0.9 - 0.2x$), solved using the neighbor-modulated fitness approach (a-c) and an exact genetic model (d-f). Grey regions indicate that the variant trait value (vertical axis) has higher fitness than the resident (horizontal axis). Lines separating grey and white regions indicate where variant and resident fitnesses are equal, and the equilibrium of the model is where these lines cross. All cases of branching point equilibria are indicated with arrows. In general, these examples show that neighbor-modulated fitness is a close approximation to fitness in the exact genetic model.
Chapter 6
A test of cooperation over resources in groups of related plants

6 Abstract

A plant’s strategy for acquiring resources should sometimes depend on information about the identity of neighbours. This study tests the hypothesis that plants cooperate (show restraint in competition) with genealogical relatives but compete intensely with non-relatives. In a greenhouse experiment with *Ipomoea hederacea*, sibling groups from the same inbred line were relatively uniform in height, whereas groups of mixed lines were often dominated by the tallest plant from the most productive genotype. Averaged across all height classes, mixed and sibling groups yielded a similar number of seeds, implying no difference in the intensity of competition. Yet sibling groups from the tallest height classes yielded significantly more seeds than mixed groups of the same mean height. These highly variable mixed groups also had significantly larger roots than sibling groups of similar mean height, and overall, siblings produced more seeds per unit mass of root. Hence, although this study does not lend outright support for the cooperation hypothesis, it is the first to link an apparently cooperative behaviour (relatively restrained root growth within productive sibling groups) with an increase in the mean fitness of related plants.

6.1 Introduction

Most plants live in social groups, where neighbours interact strongly in competition for above- and below-ground resources. Furthermore, because of limited dispersal, neighbouring plants are often close genealogical relatives (Vekemans & Hardy 2004). The effect of relatedness on the intensity of intraspecific competition could inform two important perspectives in evolutionary ecology. One common view is that a group of related (and thus phenotypically correlated) plants will inevitably suffer more competition than a diverse group that is better able to partition resources (e.g., Ellstrand & Antonovics 1985). This hypothesis plays a key role in theory for the

---

5 Unpublished manuscript at time of thesis submission.
The evolution of sex (Maynard Smith 1978; Young 1981) and outcrossing (Schmitt & Ehrhardt 1987) and has received some support in plant ecology (Antonovics 1978; Cheplick & Kane 2004; Milla et al. 2009 and refs. therein). The alternative perspective views a plant’s access to resources as a more sophisticated allocation strategy (Novoplansky 2009): further allocation to competition gives an individual a larger share of contested resources, but competitiveness is at the expense of reproduction. In this case, a group of plants with correlated strategies could avoid competition (and maximize productivity) by restraining allocation to competitive traits such as height (see Jennings & de Jesus 1968) and root growth (see Maina et al. 2002; O'Brien et al. 2005).

Plant competition may therefore exemplify a ‘tragedy of the commons’ (reviewed by Rankin et al. 2007), where escalating within-group competition acts to diminish group efficiency. Individual competitiveness is adaptive when neighbours are unrelated because a strategy of cooperative restraint is open to exploitation by more selfish neighbours (e.g., Gersani et al. 2001; Zhang et al. 1999). If relatedness is high, however, an individual can do best by cooperating because its neighbours will tend to show similar restraint (Frank 1998). The best examples of apparent cooperation among plants come from agriculture, where weak competitors have been artificially selected to maximize the collective yield of crops in monoculture (see Donald 1968; Denison et al. 2003). But given the implications for plant ecology and evolution, it is of great interest to uncover examples of cooperation that can be attributed to selection in plants’ natural social environment (see Callaway & Mahall 2007; de Kroon 2007).

One effective way for plants to coordinate resource use is to assess the relatedness of neighbours and preferentially cooperate with kin over non-kin. In the extreme, an individual organ or module should recognize and avoid competition with other parts of the same plant, and indeed, such self-recognition and avoidance seems common (reviewed by Novoplansky 2009; but see Hess & Kroon 2007). More uncertain is whether plants recognize and coordinate their behavior with related but physiologically separate individuals (e.g., siblings). The few studies that demonstrate higher productivity in related versus non-related groups (e.g., Andalo et al. 2001; Donohue 2003; Tonsor 1989) do not actually imply a cooperative strategy, and a recent study that reports restrained root growth within kin groups (Dudley & File 2007) did not measure the reproductive success of those groups. Measuring fitness is critical for any test of a putative adaptation, but it is particularly important in this context because reduced growth can imply
increased, rather than relaxed, competition. Furthermore, most studies of coordinated plant behaviour have focused on root growth, so it is timely to explore other potentially cooperative traits. Allocation to height is especially appealing because the asymmetric nature of light competition (where the tallest plants monopolize light; Weiner 1990) means that a group’s productivity can certainly depend on coordinated similarity.

Using groups of the ivy-leaf morning glory (*Ipomoea hederacea*), the current study measures both below- and above-ground responses, including lifetime reproductive success, to a manipulation of relatedness and light competition. Morning glory, a vine with gravity-dispersed seeds, is appropriate for a study of plasticity in competitive behaviour because its evolutionary history likely includes competition with siblings (when maternal plants grow vertically) and non-siblings (when maternal plants grow laterally). If plants can restrict allocation to potentially costly competitive traits, then sibling groups in competitive contexts are expected to grow to shorter heights and/or grow smaller roots than non-sib groups in the same context. Accordingly, the cooperation hypothesis predicts higher mean fitness in sibling groups; the alternative, the resource partitioning hypothesis, predicts higher mean fitness in non-sibling groups.

6.2 Methods

6.2.1 Greenhouse experiment

The experiment was performed at the University of Toronto, where plants were grown from seed in September of 2008 until January of 2009. All *I. hederacea* seeds used in the experiment were the progeny of one of three inbred lines, herein referred to as blue, red, and black. The lines were originally collected from adjacent populations in North Carolina, U.S.A. and then self-fertilized and bred by single-seed descent for three generations in a common greenhouse environment (by J.R. Stinchcombe, U of Toronto). Seeds from the same inbred line are therefore siblings with a high level of genetic similarity. Seeds were sown at constant density in pots of either three siblings or three non-siblings (one seed from each of the three lines; herein ‘mixed’ pots). All pots were the same size (12.7 cm circumference) and contained the same soil medium (80% Promix BX, 20 % Profile ceramic conditioner, and Osmocote fertilizer at 4ml/L).
On each of seven tables in the greenhouse, 15 pots were uniformly arranged (at least 18 cm apart) and assigned to a relatedness treatment in a randomized block design (where each table was a block). Pots were assigned so that on every table each of the sibling pot types (blue, red, and black) appeared at least twice. The second factor, shoot competition, was assigned in an alternating fashion: every second pot location on a table was assigned to ‘high shoot competition’ so that directly adjacent locations were ‘low shoot competition’. High competition pots had a single wooden stake fixed next to the pot on which all three plants entwined; low shoot competition pots had three stakes, one for each plant, spaced 16 cm apart. The purpose of this treatment was to give high shoot competition plants a stronger signal of light competition (via shading) than plants in the low competition group (see Weiner 1986).

Throughout the growing period, lighting and day length depended on the ambient conditions, although supplemental lighting (5 hours midday) was provided during germination and fruiting stages. Temperature was maintained at 22 °C, and all pots were watered similarly. By the end of the experiment (accounting for the loss of some pots, missing completely at random), there were 52 mixed groups (24 low and 28 high shoot competition) and 47 sibling groups (23 low and 24 high shoot competition).

6.2.2 Data collection and analysis

Plants flowered and autonomously self-fertilized in the greenhouse; later, as plants senesced, the fruits of each plant were collected. Each vine was then clipped at soil level, and its (uncoiled) height was measured. In cases where plants had multiple vines, the total plant height (a proxy for plant size) was also measured. Roots of the three plants per pot could not be easily untangled, so the combined root mass was measured. I measured roots from a particular sub-set of pots: 21 mixed groups were paired with a sibling group (11 blue, 10 red) that produced a similar total plant height (within 5 cm). This sampling design was meant to pair groups of plants that were in similar overall condition. Roots were carefully washed of all soil, dried in a forced convection oven (30 °C for 120 hrs), and weighed.

All analyses were performed with JMP 5.01 (SAS Institute). Because the pot was the experimental unit, all data were first summarized as group means or an alternative measure at the pot level. Above-ground data were treated as arising from a randomized complete block design, with Table as a random blocking factor. Root mass data were treated similarly, with
Pair as the random blocking factor. All models with random factors were analyzed with the REML method. Throughout the results, means or least squares (LS) means are presented as: estimate (lower 95% confidence value, upper value).

6.3 Results

6.3.1 Plant size and within-group size inequality

If related plants were to restrict allocation to growing tall in competition for light, then sibling pots in high shoot competition should have: (1) shorter average height than mixed pots in high shoot competition, and (2) similar average height to both low shoot competition groups. Hence, this prediction translates to an interaction between Relatedness and Shoot Competition for measures of plant height. However, there was no evidence for such an interaction on the height of the tallest vine in a pot, and in general, the shoot competition treatment had no effect on plant size (Table 6.1). The only clear treatment effects on plant size were: (1) the tallest vines and overall largest individuals tended to be in mixed pots (Table 6.1 A and B, respectively); and (2) variation in total height within mixed pots was much greater than within sibling pots (Table 6.1 C). Specifically, as mean plant size increased, mixed pots tended to be dominated by the largest plant in the group (the blue genotype, in all but four cases); in sibling groups, however, size variation increased relatively little with increasing mean size (Fig. 6.1).

6.3.2 Mean fitness of groups

There was no evidence for treatment effects when considering Relatedness, Shoot Competition, and their interaction as predictors of mean fitness (measured as seed yield) (all $P$-values > 0.8). At the individual level, plants from the blue genotype had highest yield when dominating mixtures, plants from the red genotype had highest yield when with siblings, and plants from the black genotype were relatively invariant in yield (Fig. 6.2). Altogether, these patterns resulted in roughly equal mean yield between relatedness treatments (Fig. 6.2).

The best predictor of mean fitness was the mean total height of plants in a group (see Table 6.2 A), indicating that plant size could be a useful covariate. It was particularly informative to consider Relatedness in combination with mean height and a second covariate: the variability in
height within groups, measured as the coefficient of variation (CV = standard deviation/mean). With the mean and CV in total height included in the model, sibling groups had significantly higher LS mean yield than mixed groups (Table 6.2 A, where LS means are adjusted to the mean value of the covariates; Milliken & Johnson 2002). Most importantly, however, there was strong evidence for interactions between both covariates and the relatedness treatment (Table 6.2 A). Interaction effects were explored by fitting two regression planes, one for each treatment level, from a multiple regression of yield on both covariates. The regression planes indicate that sibling groups from intermediate to large size classes in particular tended to have higher yield than mixed groups of similar mean height but greater size variation (Fig. 6.3).

6.3.3 Root system size

If the relatively low LS mean yield of mixed groups was a consequence of increased allocation to below-ground competition, then mixed groups might have larger root systems than sibling groups of the same mean height. Overall, the mean difference between mixed and sibling groups (paired to match in mean height) was positive but not significantly greater than zero ($t_{20} = 1.60; P = 0.12$; Fig. 6.4 B). However, it was the mixed groups with greater size variability than their paired sibling group (filled circles in Fig. 6.4 B) that tended to have relatively large roots. Among these pairs in particular, the average mass of roots in mixed groups was indeed significantly greater than in sibling groups ($t_{15} = 2.66; P = 0.0179$). One potential explanation for the large root masses in mixed groups is that the typically short plants from the black genotype might have had especially large roots (which may have been evident in their [unmeasured] sibling groups). This seems unlikely, however, given a significantly positive linear relation between root mass and total plant height in the groups that were measured (data not shown).

Finally, even though some sibling groups had larger roots than their paired mixed group, the average sibling group still produced significantly more seeds/mg of root (Fig. 6.5 A; $t_{20} = 2.58; P = 0.0179$). To determine the independent effect of root mass on a group’s mean yield, I considered an ANCOVA model with root mass included as a predictor (Table 2 B). The main result was an interaction between root mass and relatedness, showing that mean yield increased significantly with root mass in sibling groups but not in mixed groups (Fig. 6.5 B).
6.4 Discussion

Given that the overall mean yields of sibling and mixed groups were equivalent, the current study cannot definitively reject the a priori predictions of either the cooperation hypothesis or the resource partitioning hypothesis. If non-siblings were better able to partition resources, then the mean fitness of mixed groups should have exceeded the mean fitness of sibling groups. More stringent versions of the resource partitioning hypothesis require that the mean yield of mixed groups exceed the mean yield of the most productive monoculture (here, siblings from the blue genotype) (e.g., Trenbath 1974; Vandermeer 1989). Clearly, however, the current study provides no evidence for such a degree of mixture ‘overyielding’—a result that is consistent with the majority of studies on the productivity of mixtures (reviews by Trenbath 1974; Bell 2008).

In fact, when comparing only the tallest groups of morning glory (presumably the groups that were in best condition), it was the sibling groups that consistently overyielded the mixtures. The most productive mixed groups were highly variable in size and tended to have larger roots and lower mean fitness than the most productive sibling groups. At first glance, these patterns appear consistent with the cooperation hypothesis: the most productive mixed groups seem to suffer a group-level ‘tragedy’ of competition that is somehow alleviated within the most productive sibling groups. This interpretation implies a fascinating sophistication to plants’ competitive behaviour, but it also demands a careful analysis. The remainder of the discussion will weigh the evidence for the cooperation hypothesis, and in doing so, will identify areas for future investigation.

First, it is important to recognize that the large size inequality within productive mixed groups could result from different genotypes simply having different average growth rates (Tonsor 1989) and hence different competitive abilities in mixture. A related complication is that individuals from a given genotype were either more or less productive in mixed groups, depending on the genotype’s competitive ability in mixture (see also Masclaux et al. 2009). Nevertheless, a critical test of a group-level tragedy of the commons (TOC) is to determine whether the relatively low mean fitness of the most productive mixed groups reflects a shift in allocation to competition at the expense of reproduction (e.g., Maina et al. 2002; OBrien et al. 2005).
This test is particularly critical because size inequality alone can lead to reduced (or increased) mean productivity, even in the absence of sophisticated allocation strategies. For example, Tonsor (1989; see also Donohue 2003) suggested that if small plants suffer a disproportionate reduction in reproductive success (e.g., if they fall below a threshold reproductive size), then, as long as the mean plant size is above the threshold, genetically-diverse groups that vary in size are likely to have lower mean fitness than uniform groups. Laird & Aarssen (2005) have since generalized this phenomenon to any case in which the relationship between plant size and productivity is diminishing (an example of Jensen’s inequality). In the Appendix, however, I show that in the current study, the relationship between individual plant size and reproductive success was very close to linear. This suggests that the difference in mean fitness between the most productive mixed and sibling groups was not just an inevitable consequence of size inequality.

Furthermore, evidence suggests that the observed differences in mean fitness may indeed correspond to shifts in plants’ allocation of resources. Although there was no indication of differential allocation to plant height in response to the shoot competition treatment, mixed groups with the greatest size inequality allocated more biomass to roots and produced fewer total seeds, on average, than sibling groups of the same mean height. Moreover, whereas each additional unit of root mass translated into a significant increase the mean yield of siblings, the same increase in the root mass of mixed groups had little effect on yield. Together, these results are consistent with the expected inefficiency of escalated competition among non-relatives and the expected efficiency of cooperation among relatives.

Similar evidence of a difference in the mean size of roots between kin and non-kin groups has been interpreted as evidence of kin recognition in plants (Dudley & File 2007). However, the kin recognition interpretation has so far relied on the presumption that lesser root growth in kin groups is a cooperative strategy (i.e., adaptive because of fitness benefits provided to neighbours; see West et al. 2007). The current study is the first to link a predicted change in root growth with a correlated increase in the mean fitness of related plants, thus addressing a major shortfall in evidence for below-ground cooperation among kin (discussed by de Kroon 2007; Callaway & Mahall 2007). Still, the mechanism of recognition, or whether true recognition occurs at all, remains unknown. As noted earlier, similar evidence for cooperation among roots exists for plants growing in self and non-self contexts, where in the former context,
plants appear to gain a fitness advantage by limiting root growth. A common objection to this evidence relates to a potentially confounding effect of pot size that is relevant to most existing studies of the TOC in root growth (see Hess and de Kroon 1997). It is important to note that the objection does not apply to the current study, which used only a single pot size.

One limitation of the methods here (and in Dudley & File [2007]) is that root system size could only be measured at the group level. Hence, it is unclear how the measured traits of a group reflect the potentially adaptive allocation strategies of individual plants. Nevertheless, results from the current study suggest some interesting prospects for future research. Plants may have confrontation and avoidance strategies that depend on (unknown) cues of relatedness in a similar way that plants respond to self and non-self interactions (reviewed by Novoplansky 2009). It is particularly intriguing that although the most productive sibling groups of morning glory had smaller root systems than mixed groups on average, they also produced more seeds per unit mass of root, especially in groups with relatively large roots. This suggests that the interacting roots of relatives might have mechanisms that not only limit further growth (cf. Falik et al. 2003) but also coordinate segregation (cf. Holzapfel & Alpert 2003).

Overall, the results from the current study highlight important difficulties in understanding the nature of kin and non-kin competition in experimental settings. Particularly inconsistent results have emerged from studies that measure the performance of ‘target’ individuals within kin and non-kin groups (e.g., Argyres & Schmitt 1992; Cheplick & Kane 2004; Masclaux et al. 2009). But given that genotypes often differ in competitive ability, it seems inevitable that targets from a particular genotype will perform better or worse with kin, depending on the genotype’s competitive ability in mixture (as in the current study; see also Masclaux et al. 2009). An alternative is to focus on the mean productivity of groups, but with this approach too, it is usually unclear why kin groups sometimes perform better (e.g., Tonsor 1989; Donohue 2003) and sometimes worse (e.g., Milla et al. 2009). Here, I show that only the most productive sibling groups had higher mean fitness than the most productive mixed groups. This evidence could certainly be strengthened by studying additional genotypes, perhaps grown at multiple densities. Still, the current study is the first to link an increase in the success of related plants with a potentially cooperative allocation strategy. It would be useful if future studies also focused on both the outcome of kin and non-kin interactions (in terms of mean fitness) and the potential allocation strategies involved. Finally, it should be clear that not all plant species will
have the ability to recognize and cooperate with relatives. Beyond the conditions already discussed, it is also necessary that plants minimize competition among related seeds (see West et al. 2002) by exporting the fruits of cooperation in space or in time (e.g., into a local seed bank, as might be the case with *I. hederacea*). Hence, future tests of cooperation among plants must first consider the potentially complex social ecology of the species of interest.
6.5 Literature Cited


Donald, C. 1968 The breeding of crop ideotypes. Euphytica 17, 385-403.


6.6 Appendix

6.6.1 The relationship between individual plant size and reproductive success

Figure 6.A.1

The relationship between individual plant size (total plant height) and total seed production (yield). Data from all plants in the current study are plotted in A, whereas B-D separate the data into plant genotypes. As in the main text, the color of the circles corresponds to genotype, open circles correspond to plants in mixed groups, and filled circles correspond to plants in sibling groups. Least squares regression lines are included as a reference.
6.7 Figures

Figure 6.1

Relationship between mean total height and absolute variation (standard deviation; [A]) or normalized variation (coefficient of variation; [B]) in the total height of plants in sibling groups (filled circles; color corresponds to genotype) and mixed groups (open circles). Estimated slopes in A: sibling groups: 0.19 (95% confidence interval: 0.10, 0.30); mixed groups: 0.77 (0.60, 0.94). Estimated slopes in B: siblings groups: 0.0032 (0.0013, 0.0051); mixed groups: -0.0002 (-0.001, 0.0009).
Figure 6.2

Estimated mean yield of mixed and sibling groups (from raw data) with 95% confidence intervals given by the vertical span of the mean diamonds. The coloured lines correspond to genotype, where the endpoint of each line is the mean yield of individual plants from the corresponding genotype in mixed and sibling groups.
Figure 6.3

Predicted mean seed production (yield) for plants in sibling groups (filled circles; color corresponds to genotype) and mixed groups (open circles). The predicted values come from separate multiple regression models (one for sibling groups, one for mixed groups) that included Table (a random blocking factor), the coefficient of variation (CV) in total height (x-axis), and the mean total height (z-axis). The points shown fall upon the least squares regression plane predicted by the corresponding model. Estimated slopes: mixed groups: -3.5 ($P = 0.47$) in the x direction, 0.36 ($P < 0.0001$) in the z direction; sibling groups: 24.1 ($P = 0.012$) in the x direction, 0.50 ($P < 0.0001$) in the z direction. Both models explained a similarly large proportion of variation in the data (for both full models, $R^2 \approx 0.85; P < 0.0001$).
Figure 6.4

(A). Relationship between mean total height and the coefficient of variation (CV) in total height of plants in the sibling groups (filled circles; color corresponds to genotype) and mixed groups (open circles) that were used for root measures. The groups were paired (one from each treatment) to match in mean total height. (B). The difference in root mass between paired mixed and sibling groups (y-axis) plotted against the difference in the CV in total height for each pair (x-axis). A regression of y on x gives a positive slope of 119.1 (95 % confidence interval: 45.4, 192.8). The black lines indicate the estimated mean difference in root mass for all pairs (dashed lines show 95% confidence); grey lines indicate the mean difference when using only the pairs in which mixed groups had the larger CV (values > 0 on the x-axis; filled circles).
Figure 6.5

(A). Mean seeds per unit mass of root produced by mixed and sibling groups, where the vertical span of each diamond is a 95% confidence interval. The mean values are centered to account for variation among pairs (blocks). (B). The partial effect of root mass on the mean yield of sibling groups (filled circles; color corresponds to genotype) and mixed groups (open circles). Residual values (y-axis) account for all other variables of the model in Table 6.2 C. Estimated slopes: sibling groups: 0.11 (95% confidence interval: 0.048, 0.17); mixed groups: 0.019 (-0.039, 0.078).
6.8 Tables

Table 6.1

Results from ANOVA models that predict above-ground growth responses to shoot competition and relatedness treatments. All models also include Table as a random blocking factor (not shown). Effect sizes (with 95% confidence intervals) give the mean difference in least squares means between treatment levels, as estimated by a Tukey test.

<table>
<thead>
<tr>
<th>Response variable factors</th>
<th>F_{1,90}</th>
<th>P-value</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Height of tallest vine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shoot competition</td>
<td>0.22</td>
<td>0.64</td>
<td>low&gt;high; 2.5 cm ( -8.1, 13.0)</td>
</tr>
<tr>
<td>relatedness</td>
<td>4.0</td>
<td>0.049</td>
<td>mix&gt;sibs; 10.6 cm (0.07, 21.2)</td>
</tr>
<tr>
<td>shoot comp x relatedness</td>
<td>0.039</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>B. Total height of largest plant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shoot competition</td>
<td>0.082</td>
<td>0.78</td>
<td>high&gt;low; 2.3 cm ( -18.8, 14.2)</td>
</tr>
<tr>
<td>relatedness</td>
<td>12.4</td>
<td>&lt;0.0001</td>
<td>mix&gt;sib; 29.1 cm (12.6, 45.5)</td>
</tr>
<tr>
<td>shoot comp x relatedness</td>
<td>0.091</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>C. Std. dev. in total height</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shoot competition</td>
<td>0.23</td>
<td>0.63</td>
<td>high&lt;low; 1.8 cm ( -5.8, 9.4)</td>
</tr>
<tr>
<td>relatedness</td>
<td>28.2</td>
<td>&lt;0.0001</td>
<td>mix&gt;sibs; 21.3 cm (12.7, 27.8)</td>
</tr>
<tr>
<td>shoot comp x relatedness</td>
<td>0.22</td>
<td>0.64</td>
<td></td>
</tr>
</tbody>
</table>
Table 6.2

Results from ANCOVA models that predict a group’s mean seed yield. The model in A also includes ‘Table’ as a random blocking factor, and B includes the random factor ‘Pair’. Effect sizes (with 95% confidence intervals) give the mean difference in least squares means between levels of relatedness (estimated by Tukey tests) or, for covariates, a linear regression coefficient.

<table>
<thead>
<tr>
<th>Response variable factors</th>
<th>$F_{1,90}$</th>
<th>$P$-value</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Mean yield</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relatedness</td>
<td>13.4</td>
<td><strong>0.0004</strong></td>
<td>sibs$&gt;$mix; 5.20 seeds (2.37, 8.03)</td>
</tr>
<tr>
<td>mean total height</td>
<td>250.5</td>
<td><strong>&lt;0.0001</strong></td>
<td>0.43 (0.38, 0.48)</td>
</tr>
<tr>
<td>relatedness x mean height</td>
<td>7.42</td>
<td><strong>0.0078</strong></td>
<td></td>
</tr>
<tr>
<td>CV in total height</td>
<td>20.2</td>
<td><strong>0.041</strong></td>
<td>9.62 (0.41, 18.8)</td>
</tr>
<tr>
<td>relatedness x CV</td>
<td>10.2</td>
<td><strong>0.0020</strong></td>
<td></td>
</tr>
<tr>
<td>B. Mean yield (subset)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>relatedness</td>
<td>5.28</td>
<td><strong>0.035</strong></td>
<td>sibs$&gt;$mix; 7.52 seeds (0.58, 14.5)</td>
</tr>
<tr>
<td>CV in total height</td>
<td>0.78</td>
<td>0.78</td>
<td>9.49 (-13.3, 32.3)</td>
</tr>
<tr>
<td>relatedness x CV</td>
<td>0.27</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>root mass</td>
<td>3.98</td>
<td>0.063</td>
<td>0.10 (-0.0063, 0.21)</td>
</tr>
<tr>
<td>relatedness x root mass</td>
<td>4.82</td>
<td><strong>0.043</strong></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Least squares means are adjusted to the mean of the covariates. Interaction effects from A are plotted in Fig. 6.3; the interaction from B is plotted in Fig. 6.5 B. In B, the blocking factor ‘Pair’ accounts for variation in mean total height.
Chapter 7
General Conclusions

7 General Conclusions

The chapters of this thesis have all dealt with aspects of the evolutionary ecology of traits involved in intra- and interspecific interactions. Of particular interest, of course, were the intraspecific (social) interactions within plant populations. My goal was to understand the outcome of natural selection on traits involved in acquiring resources, where the most adaptive trait for a plant to adopt depends on the traits of neighbours. In the competition over pollinators in particular, one reason for this frequency-dependence (from the plant’s perspective) is that a pollinator’s most adaptive foraging behaviour on a plant depends on the resources available at neighbouring plants. Hence, there is an interesting parallel between the two main topics of this thesis—plant allocation strategies and pollinator foraging behaviour—that illustrates some common themes in the study of complex adaptations.

In particular, it is interesting that from the perspective of both the plant and the pollinator, an individual can do best when it has information about the ‘players’ in its environment. Plants should acquire information about the likely traits of neighbours (measured by relatedness; Frank 1998), and pollinators should acquire information about the available resources in neighbouring patches. A central question that arises throughout the thesis is: how can natural selection incorporate such information into the adaptive strategies of individuals?

Depending on the timescale at which strategies should be adjusted (and on potential constraints faced by the organism), environmental information can be incorporated in two ways. First, natural selection can adjust strategies to fit the typical (average) social or non-social environment, resulting in individuals with fixed behaviours. In other words, selection can incorporate information into the genome, giving individuals the appearance of ‘knowing’ their typical environment. This was the case in Chapters 2 and 5, where plants’ allocation strategies were gradually adjusted to fit the average relatedness among interacting plants. Animal foragers might similarly have implicit information about the resources in their environment; however,
implicit information alone limits an individual from making more rapid adjustments to a variable or changing environment.

Hence, the second mechanism by which natural selection can incorporate environmental information is by giving individuals the ability to directly acquire information and adjust their behaviour accordingly. Evidence from Chapters 3 and 4 strongly suggests that foraging bees can adjust their foraging strategies based on learned information about the quality and distribution of resources experienced. Similarly, plants might adjust their allocation strategies based on information about the relatedness of neighbouring plants. Critically, this should be observed only if plants experience variation in the composition of social groups and if they have some mechanism for discriminating kin from non-kin. Yet evidence from Chapter 6 suggests the possibility that some plants do in fact have adaptations to recognize kin and adjust their competitive behaviour accordingly.

Contributions to the study of sociobotany

One of the main goals of this thesis was to use the modern theory of adaptation, extending from Hamilton’s (1964) inclusive fitness theory, to study the interacting traits of plants. Most existing approaches have recognized the game-theoretic nature of plant competition (see Introduction), but few studies fully appreciate the full spectrum of social partners that plants are likely to face, from complete strangers to identical clone-mates. This perspective is long overdue and will surely improve our understanding of plant adaptations. Here, I consider some important goals of a research programme in sociobotany and discuss how my thesis work has contributed.

The first goal of sociobotany should be to make qualitative predictions of how equilibrium trait values vary with measurable parameters from a model (Frank’s [1998] ‘comparative statics’). Chapter 2 makes such qualitative predictions about how plants’ investment in pollinator rewards should vary with the strength of asymmetric competition and with varying degrees of relatedness within patches. These predictions could be tested if, for instance, a large number of populations varied in a measure of the strength of asymmetric competition or in the extent of clonal growth within patches. Similarly, the model in Chapter 5 makes qualitative predictions about how the equilibrium sex allocation of cosexual, animal-dispersed plants should vary with the cost of fruit dispersal (perhaps determined by the type of frugivores that visit the population)
and with the size of patches (interaction groups). This model also predicts that disruptive selection on sex allocation can arise when the relation between allocation to fruits and fruit dispersal success is strongly diminishing and when the cost of fruit dispersal is high. It would be satisfying to confirm that these assumptions are realistic. The cosexual populations that exist today are presumably not under strong disruptive selection, however, so the absence of the predicted conditions would not seem to disprove the models. The real value of the model from Chapter 5 is to illuminate previously unknown conditions that can lead to the origin of gender dimorphism.

A second goal of sociobotany should be to improve on current interpretations of plant adaptations and to predict the presence of unknown ones. Chapter 2 predicts that if plants compete asymmetrically for pollinator visits, then investment in pollinator rewards can be driven to exaggerated values in the same way that sexual selection exaggerates the signals and weaponry of some animals. It would certainly be exciting to discover the first strong evidence for secondary sexual trait in plants. The best place to look might be in cases where plants use nectar production to encourage longer residence from each visiting pollinator. In this case, evidence from Chapter 3 suggests that a strong form of frequency-dependent selection can act to drive an ‘arms race’, where rare plants with relatively high nectar production always have an initial advantage. If the evolution of separate sexes in animal-dispersed plants is partly an adaptation to the ecological consequences of animal dispersal, then the local advantage hypothesis from Chapter 5 seems to provide a more satisfying explanation than the apparently unrealistic dispersal advantage hypothesis. Finally, in chapter 6, I predicted that if: (1) plants have an evolutionary history of growing with non-kin in some circumstances and with kin in others, and (2) plants have some mechanism to avoid local competition among seeds, then facultative cooperation might be discovered when plants are grown with kin. The evidence from my experiment with morning glory is at least consistent with an expected outcome of preferential cooperation with kin over non-kin.

Lastly, the study of sociobotany should aim to return insights to the general theory of social evolution. It seems likely that plants face unique circumstances that will inform particular aspects of social behaviour that would otherwise be understudied. Recent work on the social interactions among bacteria, for instance, have recently provided important insights into the
evolution of spite (see Gardner and West 2004; 2006) and the phenomenon of ‘greenbeards’ (Gardner and West 2010).

Plants’ sessile existence is one unique circumstance that provides opportunities to consider how organisms can regulate local competition. An important theme in social evolution in the last decade has been to understand when selection for cooperation among relatives will be counteracted by local density-dependent regulation (reviewed by West et al. 2002). The importance of local competition was featured in Chapter 5, but in the Appendix of Chapter 2, I intentionally avoided the issue altogether. In that simple model, I assumed that plants could grow clonally but that they always dispersed their seeds globally. This may indeed be the best situation for a group of plants that can cooperate with clone mates: clonal growth increases local relatedness without increasing local competition among seeds. However, it would be more interesting to treat plants’ investment in clonal growth and investment in seed dispersal as coevolving traits under various ecological (and social) circumstances. This might provide novel insight into the adaptive significance of clonal growth (modularity) from a social evolutionary perspective.

Another interesting feature of plant life is cosexuality, which provides the opportunity to study disruptive selection and the evolution of dimorphism in plant sex allocation. There are relatively few existing studies on the evolutionary diversification of social traits, so this is another area where new contributions from sociobotany can be made. My results from Chapter 5 confirm previous work but also provide novel examples in which selection for cooperation can inhibit and facilitate evolutionary diversification. This is a small but promising step towards a larger contribution to the theory of social evolution, which will remain one of the greatest future challenges in sociobotany.
7.1 Literature Cited


