Advancing the Causal Theory of Natural Selection

by

Wesley Anderson

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2016 by the
Graduate Supervisory Committee:

Brad Armendt, Chair
Richard Creath
Bruce Glymour
Ann Kinzig
Charles Perrings

ARIZONA STATE UNIVERSITY

May 2016

ABSTRACT

The Modern Synthesis embodies a theory of natural selection where selection is to be fundamentally understood in terms of measures of fitness and the covariance of reproductive success and trait or character variables. Whether made explicit or left implicit, the notion that selection requires that some trait variable cause reproductive success has been deemphasized in our modern understanding of exactly what selection amounts to. The dissertation seeks to advance a theory of natural selection that is fundamentally causal. By focusing on the causal nature of natural selection (rather than on fitness or statistical formulae), certain conceptual and methodological problems are seen in a new, clarifying light and avenues toward new, interesting solutions to those problems are illustrated. First, the dissertation offers an update to explicitly causal theories of when exactly a trait counts as an adaptation upon fixation in a population and draws out theoretical and practical implications for evolutionary biology. Second, I examine a case of a novel character that evolves by niche construction and argue that it evolves by selection for it and consider implications for understanding adaptations and drift. The third contribution of the dissertation is an argument for the importance of defining group selection causally and an argument against model pluralism in the levels of selection debate. Fourth, the dissertation makes a methodological contribution. I offer the first steps toward an explicitly causal methodology for inferring the causes of selection—something often required in addition to inferring the causes of reproductive success. The concluding chapter summarizes the work and discusses potential paths for future work.

ACKNOWLEDGMENTS

I am deeply indebted to my advisor Brad Armendt for taking me under his wing for five years and providing me with all sorts of insights into the discipline and a career in academia. Without his help, I would still be the same lost graduate student I was five years ago. Thank you very much! I have also been extraordinary lucky to have Bruce Glymour on my committee. He gave me a ton of feedback on all my work. Every time I sat down to write I couldn't help but think of the kinds of criticisms he was likely to have (and yet he still always had some!). Bruce's work has been highly influential in my own thinking, and I cannot think of a better external committee member. Thank you for the amount of time you put into me! I am also extremely grateful to the my other committee members, Richard Creath, Ann Kinzig, and Charles Perrings. I have been fortunate in having all of your critical and encouraging comments on my work during my time at Arizona State University. I have learned something unique from all of you, and I hope to pass on your wisdom to future students. I appreciate all of your help and patience! I would like to thank my wife and fellow philosopher, Valerie Racine, for all of her love, patience, encouragement, and support over the last few years. Without her I would have never finished this thing. She read every draft and challenged me on basically every sentence. The dissertation is much improved because of her. Thank you, thank you, thank you! I must also thank various other non-committee members who have helped me greatly with their comments and encouragement over the course of writing the dissertation. Thanks are due to Frederick Eberhardt, Nicolas Lanchier, J.J. LaTourelle, Manfred Laubichler (and the members of his lab), Jane Maienschein, and Samir Okasha.

I could not have written many parts of this dissertation without your patient aid. I received a lot of financial support while I was writing the dissertation. Thanks are due to the School of Life Sciences and the Center for Biology and Society at Arizona State University. I would also like to give a big thanks to the administrative staff. Jessica Ranney, Andrea Cottrell, Yvonne Delgado, and Wendi Simonson, you were all very helpful in reducing my stress related to non-academic affairs so that I could focus on my research as much as possible. Last but certainly not least, I would also like to thank my family. I have received so much love and support from Jennifer Anderson, Steve Anderson, Ryan Anderson, Helen Johnson, Tom Johnson, Gayle Palama, Brett Palama, and Sarah Hoffer throughout my entire life; thank you for that, and I hope that I have made all of you proud.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

INTRODUCTION

1. The Classical Fitness-Centric Theory and Its Formalism

Evolution occurs whenever there is a change in the frequency distribution of a trait or character variable within a population. The "forces" that explain evolution include natural selection, drift, migration (or gene-flow), and mutation (Sober 1984). Lewontin (1970; cf. Lewontin 1985 and Endler 1986) states that the necessary and sufficient conditions for evolution by natural selection, in particular, are three-fold:

1. Different individuals in a population have different morphologies, physiologies, or behaviors (phenotypic variation).
2. Different phenotypes have different rates of survival and reproduction in different environments (differential fitness).
3. There is a correlation between parents and offspring in the contribution of each to future generations (fitness is heritable).

Any entities in the biological hierarchy that possess these three individually necessary and jointly sufficient conditions for evolution by natural selection are "units of selection" (Lewontin 1970, 1). Brandon (1990) has largely agreed with the above recipe, however, for reasons that will become clear below, Brandon claims that the following conditions are only necessary for evolution by natural selection:

1. Variation. There is variation in phenotypic traits among members of a population.

2. Inheritance. These traits are heritable to some degree, meaning that offspring are more like their parents with respect to these traits than they are like the population mean.
3. Differential reproductive success. Different variants leave different numbers of offspring in succeeding generations.

Evolution occurs by natural selection only if a population exhibits heritable variation in fitness. Natural selection just amounts to differences in fitness within a population. The evolution of a trait frequency distribution within the population by natural selection depends on whether these differences are heritable across generations.[1]

Before turning to what exactly fitness amounts to, it is instructive first to turn to extant formalizations of evolution by natural selection. One typical way to see this is through a derivation of Price's (1970) equation. The equation is very general and can be derived in a way that shows the link between heritable variation in fitness and evolution by natural selection. My discussion here follows closely Okasha's (2006).

Consider a parental population containing $1,2,...,i,...,n$ "units of selection." All parental entities vary with respect to some trait or character variable Z. Let z_i denote the value of the trait variable of the i^{th} parent. Suppose we want to know how the actual average of the trait variable changes over generational time. That is, suppose we are interested in the evolution of Z in the population with respect to its average $\overline{Z}$. To track

[1] Given this setup, one might guess that the focus of the dissertation is on evolution by natural selection rather than other interesting topics such as the generation or the sources of variation upon which selection acts. This is true. A more complete study of the evolutionary process is well beyond the scope of this dissertation.

evolution by natural selection in this population we need to account for heritability and fitness. For now, let us focus on fitness and take it to be realized reproductive success or realized number of surviving offspring R. Let r_i denote the value of reproductive success for the i^{th} parent. Assume that reproduction is asexual and generations are discrete. Let the average reproductive success in the parental population be denoted by $\bar{R}$, and let the relative reproductive success of the i^{th} parent be denoted by $r_i' = \frac{r_i}{\bar{R}}$.

Now consider an offspring population. It consists of all the offspring of units from the parental population and none of the parents because generations are discrete. Let $\bar{Z}_O$ denote the average of the trait variable in the offspring population. Evolution requires that $\bar{Z} \neq \bar{Z}_O$. To calculate $\bar{Z}_O$, note that the offspring population is really made up of n subpopulations, some of which might be empty. Offspring subpopulation i contains all the r_i offspring of the i^{th} parent. Let $\bar{Z}_i$ denote the average of the trait values of the offspring of the i^{th} parent. Then $\bar{Z}_O = \frac{1}{n}\sum_i r_i' \bar{Z}_i$.

The quantity the traditional Pricean framework is most interested in is $\Delta\bar{Z} = \bar{Z}_O - \bar{Z}$. Given this, we can now derive one version of the Price equation. Start with the following fact:

$$\Delta\bar{Z} = \frac{1}{n}\sum_i r_i' \bar{Z}_i - \frac{1}{n}\sum_i z_i \qquad (1.1)$$

Multiplying $\bar{R}$ by both sides, we get:

$$\bar{R}\Delta\bar{Z} = \frac{1}{n}\sum\nolimits_i r_i\,\bar{Z}_i - \frac{1}{n}\sum\nolimits_i \bar{R}z_i \tag{1.2}$$

Let $\Delta\bar{Z}_i = \bar{Z}_i - z_i$, for the i^{th} parent. Then $z_i = \bar{Z}_i - \Delta\bar{Z}_i$. With substitution, then:

$$\begin{aligned}\bar{R}\Delta\bar{Z} &= \frac{1}{n}\sum\nolimits_i r_i\,\bar{Z}_i - \frac{1}{n}\sum\nolimits_i \bar{R}(\bar{Z}_i - \Delta Z_i) \\ &= \frac{1}{n}\sum\nolimits_i r_i\,\bar{Z}_i - \frac{1}{n}\sum\nolimits_i \bar{R}\bar{Z}_i + \frac{1}{n}\sum\nolimits_i \bar{R}\Delta Z_i \\ &= \frac{1}{n}\sum\nolimits_i \bar{Z}_i(r_i - \bar{R}) + \frac{1}{n}\sum\nolimits_i \bar{R}\Delta Z_i\end{aligned} \tag{1.3}$$

Because the first term is a covariance and the second term is an expectation:

$$\bar{R}\Delta\bar{Z} = \mathrm{COV}(r_i, \bar{Z}_i) + \mathrm{E}(\bar{R}\Delta Z_i) \tag{1.4}$$

Dividing both sides by $\bar{R}$:

$$\Delta\bar{Z} = \mathrm{COV}(r_i', \bar{Z}_i) + \mathrm{E}(\Delta Z_i) \tag{1.5}$$

Now note that a trait variable's (narrow) heritability h is defined as the simple regression coefficient of $\bar{Z}_i$ regressed on z_i. So, let $h = \dfrac{\mathrm{COV}(\bar{Z}_i, z_i)}{\mathrm{VAR}(z_i)}$. The regression model is:

$$\bar{Z}_i = \bar{Z}_O + hz_i + \varepsilon_{\bar{Z}_i} \tag{1.6}$$

where the last term on the right hand side of the equation is the deviation from the linear regression line. By substitution and some covariance algebra:

$$\begin{aligned}\Delta\bar{Z} &= \mathrm{COV}(r_i', \bar{Z}_O) + \mathrm{COV}(r_i', hz_i) + \mathrm{COV}(r_i', \varepsilon_{\bar{Z}_i}) + \mathrm{E}(\Delta z_i) \\ &= h\mathrm{COV}(r_i', z_i) + \mathrm{E}(\Delta z_i)\end{aligned} \tag{1.7}$$

assuming, of course, that $\mathrm{COV}(r_i', \varepsilon_{\bar{Z}_i}) = 0$. The covariance term here is often called the selection differential. Note that $\mathrm{COV}(r_i', z_i) = \mathrm{VAR}(z_i)b$, where $b = \frac{\mathrm{COV}(r_i', z_i)}{\mathrm{VAR}(z_i)}$ is the simple regression coefficient from regressing r_i' on z_i. Finally, assume that the second term in equation (1.7)—the average transmission bias—is zero such that, one the whole, each parent transmits her trait value perfectly. Then the Price equation can be written as:

$$\Delta\bar{Z} = h\mathrm{VAR}(z_i)b \qquad (1.8)$$

The term on the right hand side of the Price equation represents the change in the average of the trait variable of interest due to heritable variation with associated differences in realized reproductive success. That is, the term represents the change in the average of the trait variable that is due to natural selection $\mathrm{VAR}(z_i)b$ and the heritability h of the trait variable. Hence, the Price equation represents the formal characterization of the classical fitness-centric theory of natural selection and evolution by natural selection—what is required is a statistical association between a parental trait variable and fitness or reproductive success.

Though the concept of heritability is of interest and central to various debates (see Lewontin 1974; Pigliucci and Kaplan 2006), this dissertation focuses on natural selection. To better understand the fitness-centric theory of natural selection, we must examine the concept of fitness more carefully, as doing so is essential to understanding the traditional views of evolutionary biology since the Modern Synthesis. In this section, I assumed fitness was identical to realized reproductive success. Yet many have found this

unpalatable and for good reasons. I turn now to the discussion of the "received" propensity view of fitness.

2. The Propensity Theory of Fitness

According to the propensity theory of fitness (Brandon 1990; Mills and Beatty 1979), fitness is a disposition or propensity of an organism or type to survive and reproduce in some particular environment. That is, fitness is the level of reproductive success an organism or type would tend to produce, given an environment and some state of the population's composition or configuration with respect to other organisms or types. This propensity is typically modeled mathematically as the expected number of surviving offspring of an unit of selection or type of unit of selection. This is sharply contrasted with how I defined fitness in the previous section. The expectation over reproductive success is meant to explain the realized frequency distribution over reproductive success, which of course might deviate from what is expected. Fitness is therefore not to be identified with realized reproductive success. The work that the propensity theory of fitness is supposed to do, then, is to explain evolution by natural selection rather than evolution by chance or drift.

To see this, consider the following. In their paper, Mills and Beatty (1979) claim that two physically identical organisms in the same environment and population ought to be given identical fitness values. Yet, if one of these twins is suddenly struck by lightening before it mates and we define fitness in terms of realized reproductive success, the two organisms seem to have different fitness values after all. The surviving twin might go on to reproduce successfully whereas the dead twin ends up with exactly zero

offspring. The lightening strike case is supposed to represent a kind of chance event. Two identical organisms are expected to have the same realized reproductive success, but by chance they do not.

Alternatively, think of a case where organism 1 has a higher propensity to survive and reproduce than organism 2 in some specified environment and population. If 1 is struck by lightening and 2 goes on to successfully reproduce, then this should intuitively be a result of drift and not selection for organisms of the type 2. The fact that 1 is fitter than 2 entails (on the propensity theory) that 1 probably will have higher reproductive success than 2. This is just the principle of natural selection, according to Brandon (1990). However, the evolutionary outcome involving a lightening strike went exactly against all expectations. Hence, in order to make sense of the distinction between evolution by selection and evolution by drift, the propensity theory is required.

Given this, it is easy to see why Lewontin's (1970) recipe for evolution by natural selection is not sufficient for producing evolutionary change by natural selection. Differential reproductive success might be the result of drift, not selection. Brandon, like other propensity theorists, attempts to distinguish between chancy outcomes and outcomes that are the result of selection by proposing that the relationship between fitness and a type in an environment-population combination is supervenient. Fitness *qua* propensity supervenes on types in their environment-population in the sense that there can be no change in fitness without a change in the relationships between the type's traits or characters and the environment-population. If reproductive success, or more precisely, the ordinal relation of realized reproductive successes among types, is not a result of the

supervening property of fitness, we do not have evolution by selection but evolution by drift.

It is useful, then, to make an explicit update to the formalism introduced above (cf. Okasha 2006). Suppose that within an environment and given some population composition or configuration, the i^{th} parent might produce 0 to m offspring. Then suppose that the parent's trait or character value along with the environment and population composition induce a probability distribution over the $m+1$ possibilities, and let w_i be the expectation of this probability distribution. The fitness of the i^{th} parent is then:

$$w_i = \sum_{r=0}^{r=m} r_i \Pr(r_i \mid z_i) \qquad (1.9)$$

Divide this by $\bar{W} = \frac{1}{n}\sum_i w_i$ to get relative fitness w_i'. Substituting into a version of the Price equation, assuming types (individuated by their trait value) breed true, gives us a new Price equation:

$$\Delta \bar{Z} = \mathrm{COV}(w_i', z_i) \qquad (1.10)$$

3. Issues with the Classical Fitness-Centric Theory of Natural Selection

Insofar as one thinks the distinction between evolution by natural selection and evolution by drift is important, the distinction made above in light of the propensity theory of fitness is helpful. Yet, the classical fitness-centric theory of selection still seems to face some challenges on this score even with the help of the propensity theory. I will first

discuss this issue before turning to a second sort of issue with the classical fitness-centric theory of natural selection.

The central case in the literature is Beatty's (1984) moth example. According to the propensity theory, there is some probability distribution over the number of surviving offspring a parent with some trait value in a certain environment-population complex is disposed to produce. Let E be a binary environmental variable, measured on parent moths, with one value for dark trees and another value for light trees. The probabilities over these values is 0.60 and 0.40, respectively. Stipulate that moths come in two shades, dark and light, and the trait variable Z measuring coloration is probabilistically independent of tree color E.[2] In other words, a dark moth has a chance of 0.60 to be found on a dark tree and a chance of 0.40 to be found on a light tree and, similarly, for a light moth. If a dark moth lands on a dark tree, suppose it will probably leave more offspring than if it lands on a light tree and conversely for a light moth because of certain facts about predation on the moth population. Beatty says "in this environment, we would say that the dark-colored moths are fitter, since the forest provides more camouflage for them than for the light moths" (1984, 194). However, suppose that the frequency with which moths land on trees of different shades diverges from probabilistic expectation, and, as a consequence, the realized average number of surviving offspring of light moths is greater than the realized average number of surviving offspring of dark moths. Here is

[2] This is in fact a standard assumption made in the philosophical literature. It rules out interesting causal structures like phenotypic plasticity, habitat selection, and niche construction. The assumption will be relaxed in later chapters.

the question: Is "the change in frequency…a matter of natural selection, or a matter of random drift?" (Beatty 1984, 195). Beatty says a bit of both. In particular, he says it is difficult to discern the distinction between unexpected results of selection *qua* differential fitness and the results of drift *qua* chancy distribution into environments.

A related case comes from Brandon (1990), but his verdict is different. Let Z and E be defined as above. Suppose that dark moths are always fitter than light moths, for whatever reason, that both types are more fit on dark trees, but light types in dark environments are fitter than dark types in light environments. One should expect again that a higher proportion of the dark moths would be the result in the next generation. However, assume that before selection the frequency distribution of individuals in environments is such that light moths are disproportionately found on dark trees and dark moths are disproportionately found on light trees so that light moths out-reproduce dark moths. Then the realized average number of offspring of light moths is greater than the realized average number of offspring of dark moths. For Brandon, if one relativizes their fitness measures to locations characterized by dark trees or relativizes their fitness measures to locations characterized by light trees, then clearly we have selection for the dark type. However, if one averages fitnesses over both locations, light moths, it was said, outcompete dark moths. In this case, says Brandon, we have evolution by drift rather than evolution by selection. The light moths, by chance, simply were more lucky in experiencing more favorable environments than the dark moths.

Given only this bit of information, it is difficult to tell who is correct and why, especially because the examples differ ever so slightly. What is agreed upon in the

literature is that the propensity theory of fitness requires some kind of theory of how to individuate a common environment-population complex. Propensity theorists must do so such that differences in fitness can be measured with respect to that common environment and, consequently, differences in fitness can be seen to explain evolution rather than haphazard distributions into multiple environments (see Shanahan 1989). Brandon (1990) famously fills this lacuna in the theory most explicitly.

Return to Brandon's example. Start by distributing the population uniformly at random over some spatiotemporal volume and divide the volume into chunks such that we can measure the fitnesses of types reliably. The ordinal relation among fitnesses defines the relevant environment—Brandon's (1990) "selective neighborhood"—as far as the theory of evolution by natural selection is concerned. Given this, it is easy to see that in Brandon's own example, there is one selective neighborhood as the dark type is fitter in all spatiotemporal chunks. However, in the example, what happens is that types are distributed into spatiotemporal locations in a biased way such that light moths enjoy higher realized reproductive success. Consequently, with respect to our selective neighborhood, what is expected to occur does not occur. Thus, Brandon says, "this is not natural selection" (1990, 46) but the result of drift *qua* the biased distribution of types into environments.

Brandon's theory of selective neighborhoods diagnoses Beatty's moth example similarly. In that case, recall, that if we were to distribute moths uniformly at random across the spatiotemporal location, light moths would be fitter in some spatiotemporal locations and dark moths would be fitter in other spatiotemporal locations. Hence, we

have two selective neighborhoods in Brandon's sense. Because Beatty calculates the fitnesses by averaging over distinct selective neighborhoods, this amounts to an analysis of the population's evolution in terms of drift rather than selection. Again, the light moth merely gets lucky in being more represented in its favorable environment than the dark moth in its more favorable environment. Though Brandon's responses to these cases have been very influential, the score is by no means settled (Gildenhuys 2009; Millstein 2002; Pfeifer 2005) and the implications of this sort of view have been greatly under-appreciated and vastly under-explored (but see Glymour 2011).

Yet another issue with the classical fitness-centric theory of selection has to do with the fact that the theory backgrounds the causal nature of natural selection. Often only implicit in the classical theory is the idea that natural selection requires that some trait or character variable causes reproductive success. However, because this causal dimension of natural selection has been deemphasized in favor of theories of fitness and statistical formulae, conceptual and methodological problems and theses arise that simply are non-issues or mistaken given clear causal reasoning. Much of the dissertation amounts to revealing these holes in the traditional framework, but to give a flavor of the issue consider the case of pure soft selection and the heated debate that surrounds it.

Assume all groups have the same number of members N. Soft selection occurs only if there is frequency (or density) dependent selection within groups and groups' (relative) fitnesses $\bar{W}'_k = \frac{1}{N}\sum_j w'_{jk}$, for all parents j in group k and for all groups k, vary little. Under experimental settings, pure soft selection occurs when

$\mathrm{VAR}(\bar{W}_k') = 0$. Note, however, that because there is frequency dependent selection within groups, the reproductive success of organisms depends on at least some aspects of the group to which they belong. That is, individual relative reproductive success r_{jk}' not only depends on z_{jk}, for all parents j in group k, but also on contextual trait variable $d_{jk} = \bar{Z}_k$, for all parents j in group k and for all groups k.

	Make-up of Group Before Selection	Group Trait	Make-up of Group After Selection	Relative Individual Fitness of Talls	Relative Individual Fitness of Shorts	Relative Group Fitness
Group 1	9 talls 1 shorts	0.10	20 talls 0 shorts	1.11	0	1
Group 2	8 talls 2 shorts	0.20	20 talls 0 shorts	1.25	0	1
Group 3	5 talls 5 shorts	0.50	15 talls 5 shorts	1.50	0.50	1
Group 4	2 talls 8 shorts	0.80	10 talls 10 shorts	2.50	0.625	1
Group 5	1 talls 9 shorts	0.90	6 talls 14 shorts	3.00	0.78	1

1.1. Pure Soft Selection.

Okasha (2004) describes the case as follows. There are 50 organisms divided into 5 groups of size 10. The trait variable measures whether an organism is tall $Z = 0$ or short $Z = 1$ and each type reproduces with perfect fidelity. The initial average value of the trait variable is 0.50. The composition of the groups and fitnesses between and within groups is described in table 1.1. The population average fitness is 2; each group doubles in size. Let us look at group 1 more closely as an example. The fitness of tall individuals

is 20/9, and their relative fitness is thus 1.11. Because short individuals' fitness is 0 the relative group fitness is $1.11 \cdot 0.90 + 0 \cdot 0.10 = 1$. By contrast, in group 5 the fitness of tall individuals is 6/1, and, hence, their relative fitness is 3. The fitness of short individuals is 14/9, and their relative fitness is consequently 0.78. Given this, the relative group fitness is again $3 \cdot 0.10 + 0.78 \cdot 0.90 = 1$.

Though each group has the same relative fitness, tall organisms get a bigger slice of the pie. There is a contextual effect on an organisms's relative reproductive success. As groups consist of more and more short organisms, all organisms have higher relative fitness. In groups 1 and 2, for example, short organisms have 0 relative fitness and talls have 1.11 and 1.25 relative fitness. Short organisms in group 5 have 0.78 relative fitness and tall organisms in group 5 have a relative fitness value of 3.

What we want to know is the change in the average value of our trait variable in the entire population due to selection. Ignoring demographic structure, we get:

$$\Delta\bar{Z} = \mathrm{COV}(w_i', z_i) = -0.21 \qquad (1.11)$$

So, $\bar{Z}$ is predicted (barring drift) to go from 0.50 to 0.29 by individual selection. Using the law of total covariance, we can derive the group selection version of the Price (1972) equation and apply it to our population:

$$\begin{aligned} \Delta\bar{Z} &= \mathrm{COV}(w_i', z_i) = \mathrm{COV}(\bar{W}_k', \bar{Z}_k) + \mathrm{E}_k(\mathrm{COV}_j(w_{jk}', z_{jk})) \\ &= 0 - 0.21 = -0.21 \end{aligned} \qquad (1.12)$$

The first term on the far right hand side is the covariance between group averages. It is an estimate of the evolutionary response of Z to group selection. The second term on the far

right hand side is the expectation of the within group covariance between individual traits and fitness. It is an estimate of the evolutionary response of Z to within group or, simply, individual selection. The result in equation (1.12) is expected. The last column in table 1 shows that there is no variation in relative fitness between groups. Both equations represent individual selection decreasing the average of the trait variable by the same amount. Yet a third statistical method called "contextual analysis" (Heisler and Damuth 1987) can be used. Contextual analysis diagnoses group selection if and only if the partial regression of individual fitness on a contextual variable (or a variable measuring the demographic conditions one experiences) is nonzero. Here we begin by regressing individual fitness on Z and contextual variable D, with values defined as above. So, start with the following regression:

$$\begin{aligned} w_i' &= 1 + \alpha_1 z_i + \alpha_2 d_i + \varepsilon_{r_i'} \\ &= 1 - 1.4 z_i + 1.4 d_i \end{aligned} \tag{1.13}$$

In equation (1.13) contextual analysis has identified an aspect of group selection, for individual reproductive success depends on the demographic conditions one finds oneself in—i.e., the partial regression coefficient α_2 that results from regressing fitness on demographic contexts, holding trait type fixed, is nonzero. Consequently, contextual analysis decomposes the evolutionary response to selection differently:

$$\begin{aligned} \Delta Z &= \mathrm{COV}(d_i, z_i)\alpha_2 + \mathrm{VAR}(z_i)\alpha_1 \\ &= 0.1 \cdot 1.4 - 0.25 \cdot 1.4 = -0.21 \end{aligned} \tag{1.14}$$

Thus we have a third predictively adequate decomposition which differs from the previous two in that it does whereas they do not diagnose the presence of group selection.

The most widespread reaction to this kind of situating in recent years has been to claim that it makes no difference which model one uses in predicting evolutionary outcomes. Indeed, as it is apparent from our three decompositions of the response to selection, whether the statistical method implies group selection (equation 1.14) or not (equation 1.12) or fails to represent demographic structure at all (equation 1.11), the predicted evolutionary outcome remains—tall types increase in frequency by 0.21. Hence, all three decompositions of the evolutionary response are mathematically equivalent—it is easy enough to express the evolutionary outcomes in either of the model types and one can use either for predictive purposes. It is in this sense that Kerr and Godfrey-Smith (2002) advocate a kind of pluralism about controversies concerning the levels of selection. Though, there may be a fact of the matter concerning whether group selection is occurring or not, either model type can be used for predictive purposes. A corollary of this view is that representing the causal structure of a population undergoing selection does not matter for predictive purposes. That is, for example, if all one is interested in is predicting evolutionary dynamics, one need not worry about representing the fact that demographic conditions cause individual reproductive success. All that is needed is some statistical decomposition of the evolutionary response to selection. Yet, as I will argue later, this model pluralism only holds given certain kinds of populations with certain causal structures over reproductive success and offspring dispersal. Hence, any pluralism about the predictive adequacy of statistical decompositions fundamentally rests

on whether certain features of populations are realized or not. Given this, it is unclear why this thesis about the equivalence of some statistical formulae is interesting.

4. The Goal of this Dissertation and Some Preliminaries

The goal of this dissertation is to advance a theory of natural selection that is fundamentally causal. By focusing on the causal nature of natural selection (rather than on fitness or statistical formulae), certain conceptual and methodological problems are seen in a new, clarifying light and avenues toward new, interesting solutions to those problems are illustrated.

Given this refocusing on what is fundamental in our theory of evolution by natural selection, Chapter 2 sets out to update explicitly causal theories of when exactly a trait counts as an adaptation upon fixation and consider implications. I do so, chiefly, by assuming the interventionist notion of causation, which, to my understanding, is the only notion of causation that is scientifically interesting as it comes with various inference procedures (Pearl 2000; Spirtes et al. 2000; Woodward 2003). A brief introduction will be useful for what follows throughout the rest of the dissertation (for those familiar with this literature, feel free to skip the next three paragraphs).

A directed graph consists of a pair $<\mathbf{S},\mathbf{A}>$, where $\mathbf{S}$ is a set of vertices and $\mathbf{A}$ is a set of directed edges or arrows. An edge is an ordered pair of vertices $<A,B>$ represented by $A \rightarrow B$. Given a directed graph, if there is a directed edge from A to B, then we say that A and B are adjacent, that A is the parent of B, and that B is a child of A. A directed path in the graph is a set of sequences of distinct vertices

$< V_1, V_2, ..., V_i, ..., V_n >$ such that for all i between 0 and n, V_i is a parent of V_{i+1}. On a directed path from A to B, A is an ancestor of B and B is a descendant of A. A path *simpliciter* on the graph is a sequence of distinct vertices such that for all i between 0 and n, V_i and V_{i+1} are adjacent. Finally, a directed acyclic graph is a directed graph where no two distinct vertices are ancestors of each other; else, it is a directed cyclic graph.

Given these conventions, a causal semantics is easily applied to directed acyclic graphs. Vertices are variables and each edge represents a direct causal relationship between pairs of variables. A causal relationship is taken to be an asymmetric relationship between variables characterizing units in a population of units. There is a direct causal relationship between variables A and B relative to some set of variables on a graph if and only if there exists at least one intervention such that changing the value of A, while holding all other variables in $\mathbf{S} \setminus B$ constant, would result in a change in the probability distribution over B. There is a directed edge from A into B if and only if the former is a direct cause of the latter, relative to the set of variables on the graph.

Furthermore, directed acyclic graphs also have a straightforward probabilistic semantics and one can translate freely back and forth between causation and probabilities given some assumptions. This will be most useful for Chapter 5, but I will introduce the basics here now. The directed graph represents joint and conditional probability distributions over the set of variables. These functions contain information on a set of conditional independence claims. From such claims, assuming the causal Markov

condition (CMC), one can infer causal relationships between variables. The CMC states that every variable is independent of its non-descendants conditional on its parents. The CMC links the probability functions with the directed graph representing causal relationships. On can go the other direction too, given the causal faithfulness condition (CFC). The probability function over the set of variables is faithful to the directed graph just in case the conditional independences implied by the probability function over the set of variables are those entailed by the CMC. Given these two assumptions, we can give Pearl's (2000) d-separation theorem. Two arbitrary variables A and B are probabilistically dependent conditional on a proper subset $\mathbf{C}$ of the set of variables if and only if there is some path on the graph between the two arbitrary variables such that no mediator— $A \rightarrow C \rightarrow B$ —or common cause— $A \leftarrow C \rightarrow B$ —in the path is in the conditioning set, and every common effect or collider— $A \rightarrow C \leftarrow B$ —on the path either is in the conditioning set or has some descendant which is in the conditioning set.

In Chapter 3, I examine a case of a novel character that evolves by (a particular kind of) niche construction and argue for interpreting this as selection for the novel character when causally understood. This interpretation implies the need to reconceptualize drift and adaptations. In Chapter 4 I analyze a population undergoing multilevel selection (on one explicitly causal view implied by some working biologists) with local frequency-dependent selection and local offspring dispersal to show that common non-causal theories of group selection force us to make bad decisions about how to represent the system, and, moreover, those representations make bad predictions. Consequently, I challenge objections against the causal understanding of group selection

and argue against model pluralism in the levels of selection debate (on the assumption that the causal theory is correct). One of the implications of Chapters 2-4 is that modeling the causes of reproductive success is necessary for understanding evolution by natural selection. In cases of interactive causation among the causes of reproductive success, we must also model the causes of selection. Chapter 5 takes some of the first steps in providing a causal methodology for inferring the causes of selection. Chapter 6 concludes by summarizing my findings and pointing toward possible avenues for future work.

Before getting to the dissertation itself, however, it is useful to briefly take a look at where the causal theory of natural selection comes from. Sober (1984) distinguishes between the "source laws" and "consequence laws" of evolutionary theory. The former describe the causes of evolution or the causes of the "forces" that enact evolution. The latter describe the evolutionary consequences of selection, drift, migration (or gene flow), and mutation, once they have been established in a population by their respective causes. Sober writes,

> Whereas it is mainly ecology that tries to provide source laws for natural selection, the consequence laws concerning natural selection are preeminently part of the province of [evolutionary] genetics. It doesn't matter to the equations in [evolutionary] genetics why a given population is characterized by a set of selection coefficients [or fitness parameters], mutation and migration rates, and so on. These values may just as well have dropped out of the sky (Sober 1984, 59).

The implication of this view is that the study of the causes of evolution and the study of the consequences of heritable differential fitness or of the covariance between fitnesses

and traits can get along quite independent of one another (cf. Mayr 1961). The consequences of differential fitness do not depend on the details of the causes of reproductive success by which fitness differences arise. Indeed, it is precisely because fitness supervenes on biological facts *cum* causal details that we can abstract over various disparate populations and see how the force of selection acts similarly across them. For example, an abstract, fitness-based analysis of all populations with heterozygote superiority will help explain the maintenance of two alleles in the population even though the underlying mechanisms may differ across such populations (Sober 1984).

However, Sober (1984) also discusses major issues with this interpretation of evolutionary theory, for which he is in large part responsible. Sober distinguishes between selection *for* and selection *of* or what is sometimes discussed as the distinction between selection and sorting. Evolution by sorting of types requires that some trait or character variable cause reproductive success. By contrast, evolution by selection for types requires that the trait or character variable that individuates the types causes reproductive success. This distinction, to my mind, is central for all causal theorists of natural selection.

Suppose that some bacteria are resistant to tetracycline and some are not. Moreover, suppose that the ones that are so resistant are blue whereas the others are red, and color does not cause reproductive success. In a tetracycline filled environment, resistant bacteria are selected for but we might also say that there is mere sorting of blue bacteria. In some sense, blue bacteria are selected. However, it is only because there is some trait variable that causes reproductive success and perfectly correlated with

coloration. By contrast, there is selection for resistance precisely because the variable which measures degrees of resistance causes reproductive success.

Causal theorists and many evolutionary biologists want to know the causal structure over reproductive success in order to know which variables take values that are selected for and which variables simply come along for the ride. Virtually every evolutionary biologist, I submit, is interested in something like the resistance variable rather than something like the coloration variable in this case. Yet this is left implicit theoretically and in practice. The classical fitness-centric theory of natural selection backgrounds the causal natural of selection. Associative measures such as covariance are bad an telling the difference between selection and mere sorting. The implications of making such distinctions with clear causal reasoning are far reaching, as this dissertation can attest.

SOME ADAPTATIONS WERE NOT POSITIVE CAUSAL FACTORS

1. Introduction[3]

Selection can relate traits to environments in broadly two ways. On the one hand, a trait can evolve as an adaptation to the total undifferentiated environment. On the other hand, a trait can also evolve as an adaptation to particular environmental conditions. The case of Endler's (1983) guppies adapting to predation conditions seems to be such an example. The two are not mutually exclusive. Presumably, an adaptation to particular environmental conditions must also be an adaptation to the rest of the environment, if it is to fix by natural selection for it. Yet the converse need not be true. An adaptation to the entire environment need not be an adaptation to anything in particular; it might simply be advantageous to its bearers no matter the environment. Adaptations driven by pure r-selection (Lennox and Wilson 1994) seem to be relevant examples. I will call those traits that are also adaptations to particular environmental conditions *adaptations to*, and I will use *adaptations simpliciter* to denote traits that are at least adaptations to the total environment and possibly, in addition, adaptations to particular environmental conditions.

In this chapter, I offer an account of adaptations *simpliciter* that accommodates all cases of adaptations *to*, *sensu* Glymour (2011). Assuming adaptations to particular environmental conditions are a proper subset of adaptations *simpliciter*, I argue that the traditional account of adaptations handed down from Sober (1984) (but see also Brandon 1990)—that the trait must have been advantageous in some contexts and disadvantageous

[3] A version of this chapter is to appear in *Philosophy of Science* (Anderson in press).

to none—is too stringent.[4] I argue for a novel theory of adaptation in its place, and then I consider the implications of my position.

Before beginning, some terminology needs to be emphasized. First, I reiterate that I assume the conception of causation from the causal modeling framework (Pearl 2000; Spirtes et al. 2000; Woodward 2003). This will aid in understanding the relationship between the position of Sober and Glymour. A trait or character variable Z and at least one environmental variable X cause a reproductive success variable R if and only if there is some set of background conditions of which $X = x$ ($Z = z$) is an element and in which changing Z (X) from $Z = z'$ ($X = x'$) to $Z = z$ ($X = x$) would change the probability distribution over R. This can be expressed graphically as: $Z \rightarrow R \leftarrow X$. The strength of these causal relationships can be, for our purposes, characterized in terms of the difference a hypothetical intervention on the exogenous causes makes to the expected reproductive success—or fitness—of organisms.

Second, in what follows it will be essential to distinguish between distinct environments. There are, in the literature, two ways of doing this. One can, as is commonly done, distinguish environments by the fitnesses of types in them (cf. Brandon 1990; but also Sober 1984). This strategy conflates locations in which different causes operate to produce the same expectations for reproductive success. The presence or absence of interactions between such environmental causes of success and phenotypic

---

[4] Though perhaps not as obvious initially, the case can be equally made for Brandon (1990). In essence, given his concept of 'selective neighborhood,' Brandon's framework diagnoses a trait as an adaptation only if Sober's does. Here I just focus on Sober's framework.

causes of success will play a crucial role in the arguments to come. Hence, I instead adopt the alternative strategy (cf. Glymour 2011) and individuate environments in terms of the values taken by variables measuring environmental causes of reproductive success.[5]

2. Sober on Adaptations

For Neo-Darwinists, selection and luck—drift in its sundry forms—are distinct sources of evolution. Adaptations are to be explained by appeals to selection for traits rather than drift. Sober seeks to define 'adaptation' so as to preserve the intuition that adaptations are driven to fixation by selection alone. So, he rules out averaging across certain environments that types happen to find themselves in by insisting that adaptations be positive causal factors for reproductive success. That is, adaptations must have increased fitness for some bearers and decreased it for none in the environments actually occupied by the species in which the adaptations arose. He defines an adaptation as follows:

> Trait z is an adaptation in population p if and only if z became prevalent in p because there was evolution by natural selection for z (cf. Sober 1984, 208).

Sober holds that there is selection for $Z = z$ if and only if individuals vary with respect to whether they have $Z = z$ and having $Z = z$ is a positive causal factor for the reproductive success of individuals in the population (1984, 280). Finally, according to Sober, a trait is

---

[5] Such variables take values that range over physically possible realizations of the represented environmental conditions. Because species' range might extend to locations that are population sinks, it is thus important that a niche be represented as a proper subset of the possible range of variable values (cf. Hutchinson 1957). Nonetheless, nothing in this chapter hinges on this, and I assume changes in variable values that generate an expected reproductive success of one for at least some types.

a positive causal factor for an outcome if and only if the trait increases the probability of the outcome for some individuals, and decreases it for none (1984, 294).

Sober employs a conception of causation on which properties are the causal relata rather than variables. Interpreting Sober's views on adaptation in the causal modeling framework used here, we have the following instead. Let $\mathbf{X}$ be a set of environmental variables each of which causes reproductive success R but none of which have a causal connection to trait variable Z nor to each other. These are the "causally relevant background conditions" (cf. Sober 1984, 289-291). Then, Sober's view requires that:

> Fixed trait $Z = z$ is an adaptation only if, for all $X \in \mathbf{X}$, Z causes R such that
>
> (S1) for some environmental condition $X = x$ and all alternative traits $Z = z'$, the fitness of organisms with $Z = z$ in $X = x$ is higher than that of organisms with any alternative $Z = z'$ and
>
> (S2) for no environmental condition $X = x$ is there some alternative $Z = z'$ such that the fitness of organisms with $Z = z$ in $X = x$ is less than that of organisms with $Z = z'$.

Constraints (S1-S2) imply that the adaptation was a positive causal factor for reproductive success. $Z = z$ is not an adaptation if, for some $X \in \mathbf{X}$, the expected reproductive success of organisms with $Z = z$ was ever lower than that of organisms with any alternative $Z = z'$ across the values of X. In such an instance, $Z = z$ will have fixed

not as the result of selection but as the result of the chancy biased distribution of types into environments in favor of $Z = z$.[6]

Sober's account of adaptation has been widely accepted, in part because it is explicitly causal, and in part because of the way in which it uses the requirement that traits be positive causal factors to avoid counterintuitive results that would otherwise follow from so called illicit averaging. In other words, it makes sense of 'common environments' in which fitness comparisons are to be conducted (see also Brandon's 1990 formulation of adaptation). Virtually everyone agrees that selection requires a common environment in which to act and the conditions encoded in (S1-S2) make good on that idea.[7] Consider an example. Suppose we have a population of spotted guppies with two different color types, one muted type with 5 bright spots ($S = 5$) and a gaudy type that has 15 bright spots ($S = 15$). Assume S has heritability 1 and there is no transmission bias. Suppose the number of offspring F is caused by the number of spots S an individual has, the number of predators $P = 0,1,2$ in the individual's surrounding, and the temperature $T = 10,20$ in degrees Celsius within the individual's surrounding, according to the equation:

---

[6] To forestall any confusion, I should note that I have just given the bare causal essentials of Sober's theory in (S1-S2). Of course, more is needed to fully represent the definition of adaptation that Sober defends. For example, the heritability of the trait variable matters here as well. However, given that I mean only to focus on what Sober requires of adaptations as it pertains to selection, I merely focus on these necessary conditions rather than all necessary and sufficient conditions.

[7] That said, serious criticisms have been made, in part on the grounds that illicit averaging is not so illicit (cf. Sterelny and Kitcher 1988).

$$\mathrm{E}(F \mid S,P,T) = 20 + S - SP + T \qquad (2.1)$$

Importantly, assume there is no causal connection between any of the causes of number of offspring so that they are all exogenous, and, further, note that, whereas spot coloration makes a contribution to the number of offspring that depends on the value of predation, there is no context-dependency between spot coloration and temperature or, for that matter, between predation and temperature. Finally, suppose that parents are replaced by offspring at random from the offspring pool with a probability that is independent of type so that the population size is constant and reproductive success R (number of surviving offspring) is a function of number of offspring F, making F a component of fitness.

Suppose that the initial population was an even mix of the two types. Assume that $T = 10$ for all guppies and mostly no predators ($P = 0$) characterizes the initial environment. Suppose that in $P = 0$ environments, the gaudy individuals stand out from their muted peers and attract more mates such that hypothetically changing from a 5 spot type to a 15 spot type would increase fitness. By contrast, under the condition of $P = 2$, the muted individuals blend in to their background better than their gaudy cousins and have a greater chance of surviving to reproductive age such that changing from a 15 spot type to a 5 spot type would increase fitness. Because P is biased toward 0 for all guppies in the initial conditions, the gaudy type probably outcompetes the muted type. This can be seen from the "absolute type expected # of offspring" column calculated as the

arithmetic mean of F given specific frequencies of 0.90, 0.05, and 0.05 for $P=0$, $P=1$, and $P=2$, respectively, as in table 2.1.[8]

	Number of Predators	Individual Expected # of Offspring	Frequency of Predation	Absolute Type Expected # of Offspring	Relative Type Expected # of Offspring
Muted 5 Spot Type	0	35	0.90	34.25	0.80
	1	30	0.05		
	2	25	0.05		
Gaudy 15 Spot Type	0	45	0.90	42.75	1
	1	30	0.05		
	2	15	0.05		

2.1. Type Expected # of Offspring for Predator Poor Environment with Temperature Set to 10 Degrees Celsius for all Guppies. Individual Expected # of Offspring is Given by Equation (2.1).

The 15 spot type is expected to have the higher fitness and, so, absent drift, it is expected to evolve. Suppose, however, that the distribution of the environmental variable P changes after some number of generations (say 5). Subsequent to this, P is distributed over S such that 5 spot types are disproportionately in $P=0$ contexts and 15 spot types are disproportionately in $P=2$ contexts. If the bias is sufficiently large, then when absolute type fitnesses are calculated by averaging over the realized environmental

[8] Properly speaking, we need to treat the within-generation fitness component as a function of the absolute arithmetic fitness component, within-generation variance in the fitness component, and population size, as in Gillespie (1974). So calculated, the fitness components for the two types in a population composed initially of 50 of each type are 34.19 and 42.23 for the muted and gaudy types, respectively. Employing the simpler and more familiar formulation changes neither the qualitative result nor the underlying intuitions. So, I employ arithmetic means for ease of exposition.

conditions, the muted 5 spot type clearly will have the advantage and will evolve if the environment does not revert. This can be estimated from table 2.2, assuming other fitness components do not cancel out what is expected to occur given expected number of offspring.[9] Assume the 5 spot type does in fact fix in the new environment. Given that a trait evolved that had the higher fitness overall, we might regard the fixed type as an adaptation (cf. Millstein 2002). However, this cannot be quite right.

	Number of Predators	Individual Expected # of Offspring	Frequency of Predation	Absolute Type Expected # of Offspring	Relative Type Expected # of Offspring
Muted 5 Spot Type	0	35	0.89	34.15	1
	1	30	0.05		
	2	25	0.06		
Gaudy 15 Spot Type	0	45	0.35	26.25	0.77
	1	30	0.05		
	2	15	0.60		

2.2. Type Expected # of offspring for an Environment with a Biased Predator Distribution and Temperature Set to 10 Degrees Celsius for all Guppies.

The envisaged fixation of the 5 spot type depends on two things: The trait variable's causal influence on the number of offspring under each level of predation, and the chancy or accidental distribution of environments, *qua* number of predators, over types. Intuitively, the frequency of the 5 spot type does not increase because there was unequivocal selection for muted types. Rather there was evolution toward the muted

[9] The Gillespie fitness components for the two types in the population are now 34.08 and 24.25 for the muted and gaudy types, respectively.

types because of the chance draw of environments for the types. The 5 spot type merely got lucky for many generations.

Sober's theory delivers the intuitively correct result exactly because he requires adaptations to have been positive causal factors for reproductive success. In the case of our toy example, the constraint (S2) is not met with respect to the environments the evolving population occupied before fixation. There were some members of the evolving population that experienced conditions under which being muted was disadvantageous. So, it is not the case that the 5 spot type is an adaptation upon fixation. It evolves merely because of the chancy distribution of types into environments in favor of the muted type.

3. Glymour on Adaptations To Particular Environmental Conditions

Sober's theory is naturally read as an account of adaptation *simpliciter*. Certainly when a trait was a positive causal factor for reproductive success, the trait is an adaptation no matter the environment. On the other hand, Glymour's view is one that (at least sometimes) diagnoses traits as adaptations to specific environmental features. Biologists are sometimes intent on showing how traits are adaptations to particular environmental variables. For example, as I mentioned before, Endler's (1983) research on guppies is paradigmatic in this respect. Other famed cases include Darwin's finches (Gibbs and Grant 1987; Price et al. 1984) and Kettlewell's (1973) moths. Recent studies on high-crowned teeth in ungulates in grassland environments are revealing as well (Heywood 2010). These studies make no sense unless they are responsive to concerns for adaptation *to*. The two conceptions of adaptations are not strictly incompatible. Yet adaptations *to* are importantly different from adaptations *simpliciter*, and sometimes Sober and

Glymour's theories deliver conflicting diagnoses on whether a trait is an adaptation or not.

To see this, let us start with what Glymour takes to be necessary for the concept of 'adaptation to':

> If we are to identify one rather than another set of features of the environment as features to which a species is adapting, those features must explain why the differences in fitness among phenotypes is one way rather than another, even when those features are homogeneously distributed over types in the population. This is only possible if the environmental features are an interactive cause of reproductive success (cf. Glymour 2011, 464).

Glymour holds that features X are homogeneously distributed over traits Z if and only if, for all values of X and all pairs of values $Z = z$ and $Z = z'$, $\text{Freq}(X = x \mid Z = z) = \text{Freq}(X = x \mid Z = z')$ (cf. 2011, 465). Moreover, though Glymour does not explicitly say so, what he means by 'interactive causation' is a special kind of interaction or context-dependency. If $Z = z$ is the supposed adaptation arising from interactive causation, then the difference in fitness made by changing from $Z = z'$ to $Z = z$ in some context $X = x$ is positive whereas the difference made by the same hypothetical intervention in some some alternative context $X = x'$ is not positive.

Glymour's view can be rendered in analogous fashion to Sober's. Again, let $\mathbf{X}$ be a set of environmental variables each of which causes reproductive success R but none

of which have a causal connection to trait variable Z nor to each other. Then, Glymour's view requires that:

> Fixed trait $Z = z$ is an adaptation to environmental features $X \in \mathbf{X}$ only if, for some $X \in \mathbf{X}$, Z and X interact to cause R such that
>
> (G1) for some environmental condition $X = x$ and all alternative traits $Z = z'$, the fitness of organisms with $Z = z$ in $X = x$ is higher than that of organisms with any alternative $Z = z'$ and
>
> (G2) for some other environmental condition $X = x'$ there is some alternative $Z = z'$ such that the fitness of organisms with $Z = z$ in $X = x'$ is not greater than that of organisms with $Z = z'$.

Constraints (G1-G2) imply that the relevant environmental conditions interacted with a trait variable in just the right way to cause reproductive success. Return to our guppy example once more to see how this works.

Suppose that predation and temperature are homogeneously distributed over types as in the initial environment (table 2.1). Homogeneity-preserving interventions on temperature can change the rate at which the population evolves but cannot change the fact that the gaudy $S = 15$ type probably fixes in the population. Increasing temperature from 10 to 20 degrees Celsius merely increases everyone's component of fitness by 10 offspring, but does not change the ordinal relation among expected number of offspring. Thus, no actual distribution of temperature can be used to explain the evolution of the gaudy type as an adaptation. If the actual distribution of temperature is homogeneous and

contrasting alternatives are too, then the same S value is expected to evolve in either case. If the actual distribution is not homogeneous, then the evolution of the trait variable is a matter of drift—the chancy assignment of one rather than another type to favorable environments—rather than a matter of the superiority of one type over the other in a common environment (cf. Glymour 2011, 466).

	Number of Predators	Individual Expected # of Offspring	Frequency of Predation	Absolute Type Expected # of Offspring	Relative Type Expected # of Offspring
Muted 5 Spot Type	0	35	0.05	25.75	1
	1	30	0.05		
	2	25	0.90		
Gaudy 15 Spot Type	0	45	0.05	17.25	0.67
	1	30	0.05		
	2	15	0.90		

2.3. Type Expected # of Offspring for Predator High Environment with Temperature Set to 10 Degrees Celsius for all Guppies.

Yet according to Glymour, predation is different. Changes in the distribution of predation can change the direction of selection even when the new distribution is homogeneous; that is, both the initial and new distribution of types to environments are such that the frequency with which 5 spot guppies find themselves in low predator environments remains equal to the frequency with which 15 spot guppies find themselves in low predation environments and so on for all values of predation. So, suppose that when predation changes, it turns out that many members of the lineage experience two predators while a few members experience one or zero predators yet P continues to be

homogeneously distributed over types. Given this and the assumption that the environment does not change again, when absolute type expected number of offspring is calculated for each type by averaging over the realized environmental conditions, the muted 5 spot type clearly has the advantage. Assuming $T = 10$ for all guppies, this can be gleaned from table 2.3.[10] In contrast to the situation described in table 2.1, the 5 spot type now has the higher fitness (again assuming no other fitness component cancels this advantage out), and it is expected to evolve in the homogeneous environment. This change has nothing to do with temperature, which increases the number of offspring of all guppies by 10 (per unit change), and everything to do with the new, yet homogeneous, distribution of predation.

Precisely because predation interacts with spot number in a special way, a change in it can influence the direction of selection, and hence the evolution of the 5 spot type is explained in part by the homogeneously distributed predation variable (cf. Glymour 2011, 466). This is because, had it been different—as in the actually realized initial environment—the ordinal relation among type fitnesses would have been different. Hence, by Glymour's reasoning, the muted type is an adaptation to the number of predators in our new environment (table 2.3).

Note, however, that on Sober's account the muted type is not an adaptation at all, because in the present case, as in the original case, condition (S2) is violated. There were some members of the evolving population that experienced conditions under which

[10] Again, for those interested, the Gillespie fitness components for the two types in the new environment are 25.69 and 16.74 for the muted and gaudy types, respectively.

having the muted trait was disadvantageous. This can be seen again in table 3 above. Hence, the fixed trait was not a positive causal factor for reproductive success in the environments actually occupied by the species in which the adaptation arose. Therefore, Glymour's theory and the positive causal factor theory of Sober conflict (at least) in this case.

To be clear, the theories do not conflict in other similar cases. If (G2) is satisfied because, but only because, there are some members that occupy environments $X = x'$ in which organisms with $Z = z$ and organisms with some alternative $Z = z'$ are equally fit, while all others occupy environments $X = x$ in which organisms with $Z = z$ are fitter than organisms with any alternative $Z = z'$, Sober's account will yield the result that $Z = z$ is an adaptation upon fixation and Glymour's account will yield the compatible result that it is also an adaptation to environmental conditions X. But, as I have argued, whenever there is any member of the population inhabiting an environment (say, $X = x'$) in which $Z = z'$ is actually fitter than $Z = z$, Glymour's account can correctly render $Z = z$ an adaptation to environmental conditions X whereas Sober's account will always, incompatibly, yield the conclusion that the $Z = z$ is not an adaptation at all.

4. A New Theory of Adaptations

The predation variable in the case of the guppies interacts with the spot variable to cause reproductive success by controlling the way in which the spot variable influences reproductive success. Under some conditions, gaudy guppies are selected for whereas under other conditions muted guppies are selected for. Endler (1983) managed to

experimentally induce a change in the direction of evolution by changing environmental variables. Nature, however, is often not as exacting as experimental set-ups. It is easy to conceive of the muted type with less bright spots going to fixation in an environment characterized by a range of predation even though it was inferior to gaudy types earlier in the lineage in environments characterized by no predation. In such circumstances, Glymour's view and the view of Sober imply inconsistent diagnoses of the trait upon its fixation in the population. Glymour's view implies that the fixed trait is an adaptation to the predation variable. Sober's view implies that the fixed trait is not an adaptation and so *a fortiori* not an adaptation to the predation variable. This is because Sober is committed to the view that adaptations were positive causal factors for reproductive success.

The correct response to this conflict, I propose, is to accept both adaptations *simpliciter* and the more restricted adaptations *to*. However, we can be pluralists in this sense only if we give up on the idea that adaptations necessarily were positive causal factors for reproductive success and, in particular, give up on constraint (S2). Though there seem to be adaptations that are not adaptations to anything in particular, it is a reasonable assumption that all adaptations *to* must also be adaptations *simpliciter*. Given this assumption, Sober cannot be right to require that adaptations were positive causal factors for reproductive success. For if we follow him in requiring this, we end up denying that certain adaptations *to* are adaptations *simpliciter*, as for example in our toy case with strictly homogeneous environments. So, contrary to Sober, in this case, the

muted type—which is an adaptation to the number of predators—must also be an adaptation *simpliciter*.[11]

As far as adaptations to particular environmental conditions go, then, let us stick with Glymour's constraints codified in (G1-G2) for now. For adaptations *simpliciter*, we must drop (S2). So, for the sake of transparency:

> Fixed trait $Z = z$ is (at least) an adaptation to the whole environment $\mathbf{X}$ only if, for all $X \in \mathbf{X}$, Z causes R such that for some environmental condition $X = x$ and any alternative trait $Z = z'$, the fitness of organisms with $Z = z$ in $X = x$ is higher than that of organisms with any alternative $Z = z'$.

This simply means that there was some environmental condition such that changing the trait variable from the alternative trait to the supposed adaptation under such conditions would have increased fitness. On the theory of adaptation on offer here, all adaptations *to* are also adaptations *simpliciter* as desired.

5. Implications

Given the centrality in biology of the idea that adaptations are sometimes adaptations to particular environmental conditions, I think we have little choice but to reject the requirement that adaptations exert a uniform positive influence on reproductive success. If we insist on it, we will not only have to give up on the idea that coloration is an adaptive response to predation, that hypsodonty is an adaptive response to some feature

---

[11] Note that, as I have argued, we do not have the option to drop Glymour's (G2) instead. For in so doing, we would not be able to make sense of either guppy type (upon its fixation) as an adaptation, let alone an adaptation to predation levels.

of grasslands, and that bill depth is an adaptive response to seed frequency. We will also have to give up on the idea that changes in environmental conditions drive evolutionary responses by means of selection within a lineage. If the homogeneously distributed frequencies over distinct environments changes and that change drives the evolution of a once rare and disfavored type, this is because the environmental variable and the trait variable interactively cause reproductive success. It will then follow that the evolved trait had a negative influence on reproductive success overall prior to the change and a positive influence overall after the change. Such traits do not have a uniform positive influence on reproductive success, and *a fortiori* cannot be adaptations unless we relinquish the requirement.

Doing so has both theoretical and practical implications. Here I examine just two. Individuating environments by vectors of values for environmental causes of reproductive success is essential for any account of adaptation *to.* It is essential for various other tasks too—e.g., accounting for phenotypic plasticity and local coadaptations—as Glymour (2011) has argued. More importantly for us, this move offers machinery for explaining where and why the evolution of a population is evolving by drift and where and why the population is evolving by selection. We have evolution by drift when we have a heterogeneous distribution of environmental causes, but on any homogeneous distribution of the environmental variables the ordinal relation between fitnesses would be reversed. It is just such cases that Sober's requirement that causes exert a uniformly positive influence is meant to exclude. Alas, the requirement goes too far, excluding also cases in which relevant environment conditions are homogeneously

distributed over types, but interact with the type variables to cause reproductive success (the paradigm case of adaptation *to* for Glymour). Still, given that environmental conditions are essentially never strictly homogeneously distributed over types, further elaborations of the theories articulated herein will be required. But the proper elaborations must tread carefully: As one weakens the assumption of homogeneous distributions, one invites illicit averaging—i.e., treating cases of drift driven evolution, in the form of the disproportionate representation of a less fit type in a more favorable environment, as cases of selection driven adaptation. A full discussion would take us well beyond the scope of this chapter (in fact, we will return to the issue in the next one), but it is sufficient here to note that any account of adaptation will need to confront this issue in one way or another, and the present account has the virtue of pressing the point directly.

The framework developed here for understanding the relation between adaptations and drift also has substantive implications concerning biological practice as well. Given the framework, it is only by modeling the causal influence of environmental variables explicitly that one can distinguish between traits driven to fixation by selection rather than by drift. The point is related to a point already made by Glymour (2011). Glymour notes that evolutionary biologists sometimes do not explicitly model environmental conditions. Instead they condition on the environment as a whole. For example, population genetics models employ a fitness parameter for each type. The parameter can be seen as an estimate of the disposition for the type to survive and reproduce in the environment taken wholly. Hence, these parameters can also be seen as an estimate of overall habitat quality from the type's point of view. Sometimes such

models can be used to diagnose when drift goes against what one would expect from selection (cf. Brandon 1990). However, such models cannot do so reliably on the view of the relationship between adaptations and drift given here.

Suppose we collect data on an individual trait or character variable and individual reproductive success for various spatiotemporal locations without measuring any environmental conditions in those locations. For each location we estimate fitnesses and assess their ordinal relations across the locations. It might be that the ordinal relation among fitnesses is distinct across some locations. We can account for this in two ways, assuming none of the environmental conditions are causally connected with the focal trait variable. Either there are some homogeneously distributed environmental conditions that interact to cause reproductive success, or there are some heterogeneously distributed environmental conditions that cause reproductive success. In the latter case, the ordinal relation among fitnesses might differ between locations without interactive causation. For example, there might be some locations with high predation for all and homogeneous temperatures fluctuations in which muted types are favored, but other locations with high predation for all but also with gaudy types being sufficiently disproportionately in warmer temperatures such that gaudy types are fitter. To distinguish between the two cases, we would actually have to explicitly model the causal influence of the environmental variables rather than merely condition on various spatiotemporal locations. Hence, the verdict that a trait is an adaptation upon fixation requires explicitly representing the causal influence of environmental variables.

6. Summary

Sober requires of adaptations that they were positive causal factors for reproductive success. Glymour's work, however, entails that fixed traits are adaptations to specific environmental features only given a special sort of interactive causation. In some cases these positions provide inconsistent diagnoses. I argued that, given one such conflict, we ought to give up on the idea that adaptations must have been positive causal factors for reproductive success. On grounds that all adaptations to specific environmental features are adaptations *simpliciter*, Sober cannot be right to require this of adaptations. I provide a rectification of the two theories, and use it to highlight two implications for evolutionary biology. Theoretically, the concepts of adaptation and drift are inseparable, insofar as an account of adaptations will need to rely on a prior decision about which distributions into environments do and which do not constitute a form of drift. Also, in practice, various attempts to measure fitness, predict the consequences of differential fitness, and identify adaptations will turn out to require quantitative measures of the environmental causes of reproductive success.

NOVELTIES EVOLVED BY NICHE CONSTRUCTION

1. Introduction

In this chapter, I am concerned to elucidate a conceptual conflict between some traditional conceptions of selection and drift, on the one hand, and the evolution of novelties by niche construction, on the other hand. Novelties are a particular kind of adaptation where the character was non-homologous upon its arrival in the population and selected for as against ancestral characters. Characters and novelties here are to be understood in the sense of Wagner (2014). Even though fins can vary within species and even though fins between species might differ functionally, fins constitute a unique character or homolog due to their identical "core" genetic basis—Wagner's "character identity networks" (cf. Wagner 2014). Given this, it is only when a change occurs in these character identity networks that we have a non-homologous character (e.g., forelimbs) that might spread through the population and fix as a novelty by selection for it (Wagner 2014).

Many proponents of the Extended Synthesis argue that niche construction is a powerful driver of the evolution of novelties (Laland et al. 2015; Laubichler and Renn 2015; Müller and Newman 2005; Sterelny 2009). If types causally influence the contexts they find themselves in, then we have niche construction.[12] Suppose also that types cause reproductive success and contexts themselves cause reproductive success. When selection

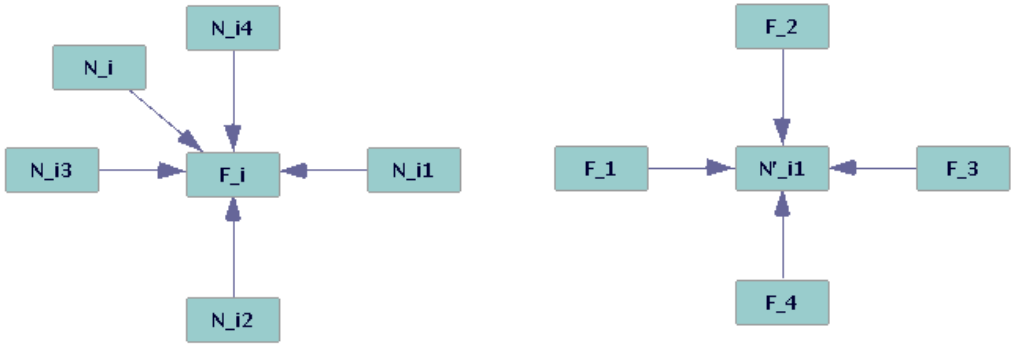
[12] The sense of niche construction used here is very broad, encompassing everything from preferential habitat choice to standard examples like beaver dams and spider webs. The example used in this chapter simply involves a kind of neighborhood selection coupled with the fact that individuals can disperse offspring only to nearest neighboring locations.

favors the evolution of novelties in those circumstances, the causal connection between types and their contexts will generate a heterogeneous distribution of types into the environment.

However, on the traditional picture, selection requires a common environment within which to operate and therefore drift in the form of the heterogeneous distribution of types into environments is a distinct process from the selection for characters within environments (Brandon 1990; Sober 1984). Consequently, in order to make sense of the evolution of novelties by niche construction as the evolution of novelties by selection for them, I argue that we must not define selection so that selection on a population requires that members of that population inhabit a common environment. I argue for this claim in two steps. First, I examine a population existing on a landscape where half of the inhabitants engage in niche construction and half do not in order to illuminate the conceptual conflict between traditional concepts of selection and drift, on the one hand, and concepts of novelties and niche construction, on the other hand. By applying those traditional concepts to the half of the population engaged in niche construction, no sense can be made of novelties evolved by niche construction as the evolution by selection for those novelties. Second, I illustrate a distinction between causal structures over reproductive success where characters are selected for and causal structures over reproductive success where characters are mere by-products of selection on something else. With that distinction, the conclusion that we must not define selection in terms of a common environment almost immediately follows.

2. The Population

Consider an asexual population with an ancestral type 0 and a non-homologous type 1. The population begins and remains at carrying capacity, with initially fewer numbers of type 1 individuals. The population exists on a landscape with a clear division that "cuts" the population in half—side a and side b—and each individual lives on a unique site or patch within their side of the landscape. On both sides generations are overlapping. Individuals are randomly and asynchronously chosen from the population and replaced. Specifically, each replacement cycle involves the following sequence. A unit is chosen and removed from the population. Then each unit within the dispersal distance of the newly vacated site reproduces with perfect heritability and without transmission bias, according to a function described below. From the pool of all offspring produced by units within the dispersal range, one is chosen at random to become the new occupant of the site, and the remaining offspring die. This sequence is repeated and sampling of the sites for updating the type of unit on them is done with replacement. Importantly, while all units reproduce offspring according to the function to be given shortly, dispersal distance varies between the two sides: on side a, the offspring can disperse only the von Neumann neighborhood of the parent whereas on side b, offspring can disperse over the whole of the side. There is some inter-mixing where the sides meet. If, for example, an individual on side b disperses to side a, that individual can do so only if the vacant site is in its von Neumann neighborhoods. If, however, an individual on side a disperses to side b, then that individual's offspring can go to the vacant site on side b, wherever that site might be on side b—that is, whether or not it is a von Neumann neighbor of the



3.1. (A) Selection: The Causal Structure for Determining Number of Offspring. (B) Dispersal: The Causal Structure for Determining the Type at the Focal Site.

individual on side a at a site where the two sides meet. I will say that dispersal on the a side is local whereas dispersal on the b side is global. In the next two paragraphs I describe the difference between the sides in more detail.

On side a the vacant location has four von Neumann neighbors—type denoted by $N_1 - N_4$. Each of these types in turn has four von Neumann neighbors including the former occupant, numbered 1-4 clockwise starting with the former occupant. That is, type denoted by N_{ij} for $\{i, j\} = \{1-4\}$ so that, for example, N_{21} is the type of the first neighbor of the second neighbor of the focal vacant spot or, otherwise put, the type of the former occupant of the focal vacant spot. Each von Neumann neighbor then generates a number of offspring; the replacement unit is drawn at random from the pool of all such offspring. The number of offspring F_i for neighbor i of a newly vacated site is

determined by a contribution from the vacated site itself, and another contribution from each of the other von Neumann neighbors, each of which are determined by the type of the unit at (or, in the case of the newly vacated site, formerly at) the site, and a final contribution determined by the type of the unit i itself. This can be represented in structural equation form as:

$$F_i = c + \alpha_1 N_{i1} + \alpha_2 N_{i2} + \alpha_3 N_{i3} + \alpha_4 N_{i4} + \alpha_5 N_i \qquad (3.1)$$

where c is a constant and there is no error. This is represented graphically in figure 3.1A. Here we will explore the case in which $c = 72$, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 67$, and $\alpha_5 = -12$. The fact that in the model the next occupant is chosen as a single survivor—the new occupant—at random from a pool formed by pooling the offspring of all the von Neumann neighbors entails both that F_i causes N'_{i1} after selection (i.e., during dispersal) and that there is only one (new) occupant. The probability that N_{i1} becomes a non-homologous type 1 individual is given by:

$$\Pr(N'_{i1} = 1 \mid F_i) = \frac{\sum_i F_i N_i}{\sum_i F_i} \qquad (3.2)$$

The probability that N_{i1} becomes an ancestral type 0 individual is given by $1 - \Pr(N'_{i1} = 1)$. The causal dependency of the new type on number of offspring is represented graphically in figure 3.1B.

On side b, by contrast, let L denote the type at the formerly occupied location and suppose the (former) occupant is a neighbor with every other individual on side b

plus some individuals on side a, where the sides meet. Each of these neighbors of the focal vacant spot, however, only has four von Neumann neighbors which might or might not include the vacant location. So suppose there are 200 individuals on this side of the population, and suppose the population is on a torus so that there are two divisions where the population sides meet and so that there are an additional 40 neighbors from side a of the focal vacant spot. As before, the number of offspring F_i for neighbor i (for $i = 1,2,...,239$) is caused by each N_{ij} (for $j = 1-4$) and its own type N_i, where N_{i1} might or might not be identical to L depending on the individual's placement across the landscape. But after the number of offspring for all neighbors $i = 1,2,...,239$ is determined, the next occupant is chosen as a single survivor—the new occupant—at random from a pool formed by pooling the offspring of all these neighbors. So, the number of offspring contributed by the i^{th} neighbor is still determined by equation (3.1) with the causal graph depicted in figure 3.1A. However, dispersal of offspring to the vacant site is no longer local, like selection, but global. Each individual gets to disperse offspring to the vacant site so that figure 3.1B should be changed to depict every F_i (for $i = 1,2,...,239$) causing L rather than merely F_i (for $i = 1-4$) causing N'_{i1}.

3. A Novelty Evolved By Niche Construction

That is our population. Given the causal parameter settings, both types on either side have a much higher expected number of offspring given more non-homologous type 1 neighbors, but being a non-homologous type 1 (rather than an ancestral type 0) comes with a cost of 12 offspring. Given this, what should we predict? Our traditional intuitions

about evolution by natural selection (and drift) have been constructed with the causal structure on side b in mind. On the traditional view, the fitness of organisms is a disposition or propensity to survive and reproduce in a specified environment. Properties of organisms generate environmentally specific dispositions to survive and reproduce so that the ordinal relation of fitnesses is stable given some set of considered types (Brandon 1990). Diagnosing a character as an adaptation requires doing so with respect to previously extant variation within a unique common environment. Only with respect to a common environment can comparing differences in fitness give us an idea of how the characters themselves (rather than something else) lead to differential fitness. To test this, ideally, we should plant individuals at sites in such a way to ensure that character type is independent of neighboring types. For otherwise, if the influence of characters on reproductive success depends (in some way) on surrounding neighboring types, then there is no way of teasing apart the difference organismal characters make to reproductive success and the difference environmental contexts make to reproductive success; and presumably, it is the former rather than the latter that drives evolution by natural selection.

In other words, like any good experimental setup, types should be scattered at random over patches at each generation. Otherwise, drift as the heterogeneous or biased distribution of types into environments can interrupt the selection for advantageous characters within a common environment. This setup is essentially what we have on isolated side b of the population. Suppose we remove side b from side a to study b in isolation. We can now think in terms of an evolutionary game context where

$F_i(1,1) = 0.25c + \alpha_1 + 0.25\alpha_5 = 82$ is the payoff of a type 1 playing against a single type 1, $F_i(0,1) = 0.25c + \alpha_1 = 85$ is the payoff of a type 0 playing against a single type 1, $F_i(0,0) = 0.25c = 18$ is the payoff of a type 0 playing against a single type 0, and $F_i(1,0) = 0.25c + 0.25\alpha_5 = 15$ is the payoff of a type 1 playing against a single type 0. A change in the frequency of the non-homologous type 1, then, can be approximated by:

$$\Delta B = B(1-B)(F_i(1,1) - F_i(0,1))B - (F_i(0,0) - F_i(0,1))(1-B)) \qquad (3.3)$$

where B is the frequency of the non-homologous type 1. Equation (3.3) is a version of the replicator equation. Given that this payoff structure amounts to a n-player Prisoner's dilemma and type 0 is the defector, replicator equation (3.3) predicts that the non-homologous type will not become prevalent (i.e., $\hat{B} = 0$). The non-homologous type, consequently, cannot become a novelty and can only become prevalent (if at all) via drift. The type 1 would have to get lucky in being on average surrounded by more type 1s than the ancestral type.

However, things are different on side a, again, taken in isolation. To get a sense of this, let us initially think of this side of the population as existing on the one dimensional integer lattice $\mathbb{Z}$. Suppose that initially there exists a finite interval of (say 6) non-homologous type 1 individuals, that the relevant neighbors are the individuals directly to the left and right of an individual, and there are infinitely many type 0 individuals. Then there are two interfaces between clusters of types. If the ancestral type next to the left interface is chosen for mortality first, then the novel type next to the left interface produces $0.5c + \alpha_1 N_{i1} + \alpha_2 N_{i2} + 0.5\alpha_5 N_i = 0.5c + \alpha_2 + 0.5\alpha_5$ and the ancestral

type next to the deceased ancestral type produces $0.5c + \alpha_1 N_{i1} + \alpha_2 N_{i2} + 0.5\alpha_5 N_i = 0.5c$. So, N_{i1} (for $i = 1,2$) changes to 1 with probability:

$$\Pr(N'_{i1} = 1) = \frac{0.5c + \alpha_2 + 0.5\alpha_5}{(0.5c + \alpha_2 + 0.5\alpha_5) + (0.5c)} \tag{3.5}$$

where parameters c and α_5 are divided by 2 because there are 2 nearest neighbors now rather than 4. Alternatively, if the leftmost non-homologous type is chosen for mortality first, then this N_{i1} changes to 0 with probability:

$$\Pr(N'_{i1} = 0) = \frac{0.5c + \alpha_1}{(0.5c + \alpha_1) + (0.5c + \alpha_1 + \alpha_2 + 0.5\alpha_5)} \tag{3.6}$$

The dynamics only depends on the magnitude of these probabilities. If the former is greater than the latter, then the interface between the types advances stochastically to the left such that all sites to the left of the interval become occupied by the non-homologous type 1 and, by symmetry, the non-homologous type fixes as a novelty. Given the causal parameter settings, this is in fact what we should predict. A type 0 occupies the space occupied by the leftmost type 1 with only probability ≈ 0.385, if that type 1 is chosen first, whereas a non-homologous type 1 occupies the space occupied by the ancestral type 0 nearest to the left interface with probability ≈ 0.73, if that type 0 is chosen first. Hence, the non-homologous type evolves into a novelty eventually (cf. Evilsizor and Lanchier 2015; Ohtsuki et al. 2006).

This result that the non-homologous type fixes as a novelty can be extended to other initial starting scenarios. Suppose that initially some finite amount of non-

homologous type 1 individuals sprout up at random across the integer lattice. In this case, there are multiple interfaces that behave like the one just examined except when both neighbors of a focal vacant space are of the same type. In such exceptions, the probability that N_{i1} changes type is equal to 1. Hence, these interfaces tend to disappear. This leaves only three special scenarios to consider where types are in clusters of size ≥ 2. I will consider them in turn.

First, there is the scenario where the interface is between three (or more) non-homologous type 1s in a row and two ancestral type 0s so that we have something like the following picture:

$$...-?-?-?-1-0-0\times 1-1-1-?-?-?-...$$

where ? can be of either type and $\times$ is the considered interface on $0.50+\mathbb{Z}$ between clusters of types. In this case, the space of the non-homologous type 1 next to the considered interface only has probability ≈ 0.385 of changing to the ancestral type 0 if chosen for mortality first whereas the space of the ancestral type 0 next to the considered interface has probability ≈ 0.485 of changing to the non-homologous type 1 if chosen for mortality first. The non-homologous type 1 stochastically wins. Second, there is the scenario where the interface is:

$$...-?-?-?-0-0-0\times 1-1-0-?-?-?-...$$

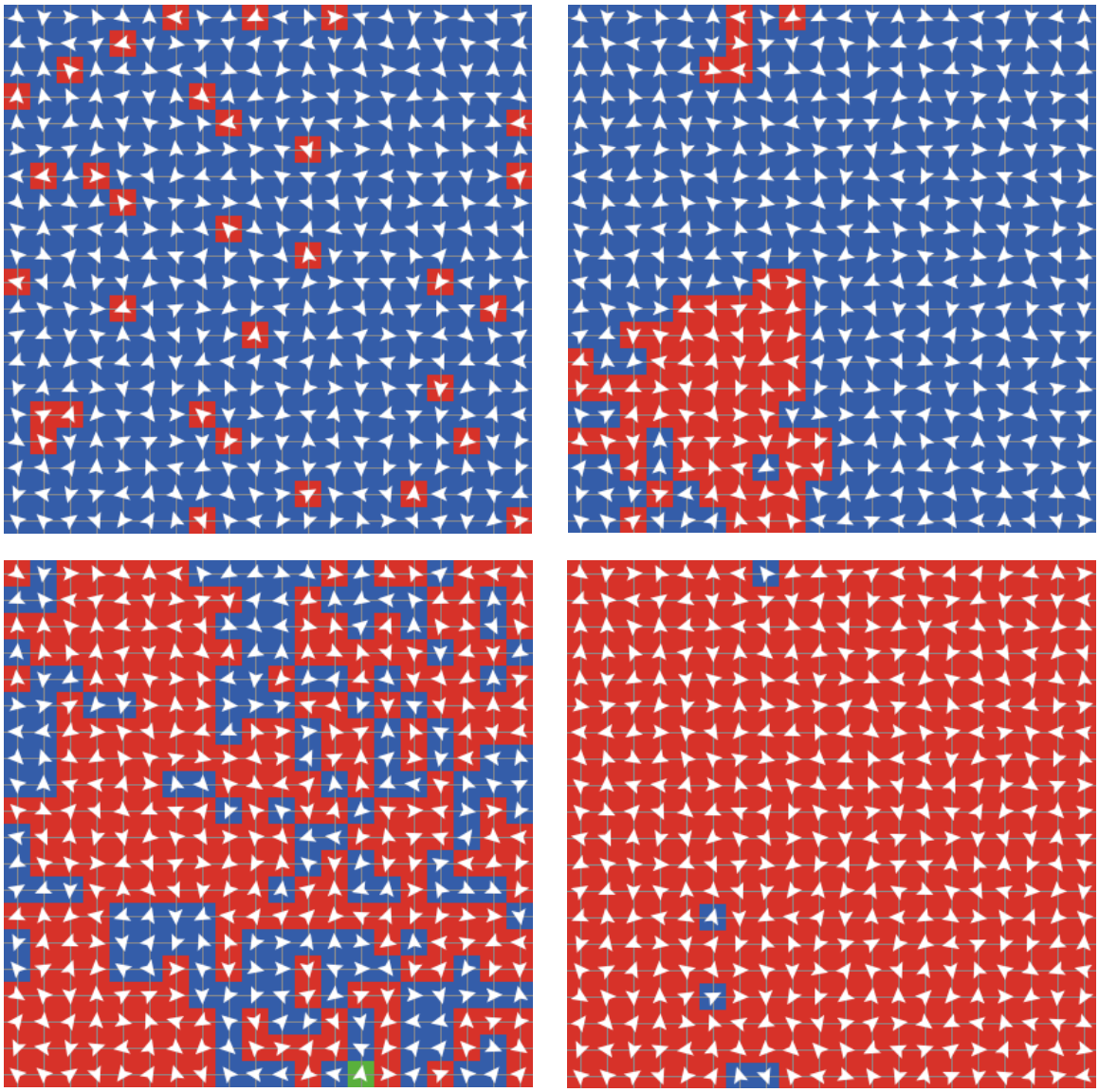
In this case, the site of the non-homologous type 1 only has probability ≈ 0.515 of changing if chosen first whereas the space of the ancestral type 0 has probability ≈ 0.73 of changing if chosen first. The non-homologous type stochastically prevails again.

Finally, there is the third scenario where the interface is such that the ancestral type 0 is predicted to prevail:

$$...-?-?-?-1-0-0\times1-1-0-?-?-?-...$$

Here, the site of the non-homologous type 1 has probability ≈ 0.515 of changing if chosen first whereas the site of the ancestral type 0 only has probability ≈ 0.485 of changing if chosen first. So the question is: What is the frequencies of these three kinds of interfaces? As Ohtsuki et al. (2006; see also Evilsizor and Lanchier 2015) have shown, the non-homologous type 1 is surrounded on average by one more non-homologous type 1 individual than is the ancestral type 0. Hence, under a general set of initial conditions the non-homologous type goes to fixation as a novelty eventually because the third scenario does not occur as often as the first two special scenarios combined.

To explore the dynamics of our original population with inter-mixing between the two sides in two dimensional space, I conducted simulations of the population existing on a torus. An ensemble of 100 populations was simulated with an average of 0.90 of 400 sites being initially occupied by the ancestral type and the rest being initially and randomly occupied by the non-homologous type. For each population, the simulations ran for 800,000 generations unless fixation occurred. As predicted by the arguments given for the one dimensional integer lattice case, the non-homologous type goes to



3.2. (A) Initial Configuration of Population 24, (B) Population 24 at Beginning of Generation 4,751, (C) Population 24 at the Beginning of Generation 12,908, and (D) Population 24 at Beginning of Generation 17,376. Type 1 is on Red Patches. Type 0 is on Blue Patches. The Unit on the Green Patch in (C) Changed from Type 1 to Type 0 in the Previous Generation.

fixation as a novelty on side a. However, once the novelty establishes itself on side a, the entire population actually ends up going to fixation for the novel type as well because there is inter-mixing between the sides. Figure 3.2C shows the population just at the mark

where it surpasses the overall frequency of the ancestral type because it creates a giant cluster of like neighbors on the left.

Irrespective of the overall frequencies of types, individuals of the non-homologous type 1 end up with more non-homologous type 1s in their neighborhood than do ancestral type 0s. This is because character type (indirectly) causes neighboring type. Consequently, non-homologous types evolve by (a kind of) niche construction quite reliably contrary to what is predicted by the replicator equation (3.4). Our intuitions have been trained by thinking of populations like side b of our population in isolation where the non-homologous type 1 is not predicted to fix by selection. We will, then, have to rethink our intuitions to avoid interpreting the novelty evolved by niche construction in our population as evolution by drift.

4. Making Causal Sense of Novelties Evolved By Niche Construction

Recall that the standard intuitions about selection are that it requires a common environment and that the heterogeneous or biased distribution of types into distinct environments is a form of drift. Further recall that, given this, the ideal experiment to sort cases of selection and cases of drift was to scatter types randomly over the landscape at each generation—severing any neighborhood effect on the influence character type has on reproductive success—so that we can arrive at estimates of type fitnesses in homogeneous environments. To save the intuition that the non-homologous character evolved by selection for it and continue holding the idea that selection requires a common environment, we might try another experiment instead. Let us scatter types randomly initially over the landscape, repeating for multiple populations, and estimate fitnesses at

each generation and spatial range by averaging over these multiple populations. The resulting fitness estimations, in effect, average over possible initial configurations and trajectories of configurations of types across the landscape. The fitnesses are what you might expect given our analysis in the previous section—the non-homologous type is favored. Yet, the fitnesses are properties of the ensemble of the possible populations rather than of the populations themselves (cf. Pence and Ramsey 2013). The generation and spatial range specific ensemble fitnesses need not correspond to the generation and spatial range specific particular population fitnesses. This is problematic insofar as one thinks models containing fitness parameters ought to represent this or that population's evolution rather than this or that type of population's evolution. Ensemble fitnesses allow us to reliably predict the arrival of the novel type but at the cost of not representing why.

Not everyone will find this troubling; some might believe it is the correct way to think about fitnesses and evolutionary models. Indeed, at least part of Sober (1984) can be read as supporting the view that fitness is usefully severed from messy causal biological facts and that in estimating fitnesses we should avoid conditioning on causes and effects of traits or characters when estimating the relationship between those characters and reproductive success. On this instantiation of the tradition, we are to think of evolutionary theory as a theory of forces. Natural selection is one such force, pushing the population through a state-space defined by type frequencies at a rate and direction in proportion to fitness parameters. Drift is another such force. Yet, this theory is not a theory of causes, representing environmental, phenotypic, and developmental causes of reproductive success. Hence, our population can be thought of in terms of selection for

the novel type at the, some would say, happy expense that our fitness parameters are divorced from the actual biological and causal details of our population.

To avoid severing the causal connection between character types and neighborhoods that make the sides of our population distinct, we might try a third experiment instead. We might scatter types in the exact configuration they were actually found in within each generation, repeat this multiple times, and estimate fitnesses by averaging over these repetitions of a generation. But here, says Brandon (cf. 1990, 60-63), we do not have selection for the novel type but rather selection for a certain kind of neighborhood—viz., one characterized by an abundance of the non-homologous type 1. This way we have selection for a certain kind of neighborhood and character type is just a by-product of this selection process.

In some sense, character type has nothing to do with the success of individuals with the non-homologous character. Rather the non-homologous character increases in frequency because they more commonly experience the non-homologous type 1 neighborhood. There is selection within a common environment, but it is in favor of a non-homologous type 1 neighborhood that non-homologous type 1 individuals produce rather than in favor of the non-homologous type 1 character itself. On this instantiation of the traditional framework, no sense can be made of the idea that the novelty arises because it was selected for in our population. It free-rides on the benefit bequeathed by the non-homologous type 1 neighborhood.

I do not want to claim here that what the above traditional frameworks imply about our population is necessarily incorrect. Yet, I suspect that many will find the above

interpretations lacking. Doubtless many will not think of it as a case of evolution by drift nor as a case of selection acting on an ensemble of populations. Indeed, advocates of the Extended Synthesis will be troubled by the call to average over possible initial configurations and trajectories of configurations of types as doing so fails to represent how niche construction can be a powerful driver in the evolution of novelties. In addition, I think the the intuitions that the novelty arises as a by-product are mistaken because they conflate distinct kinds of causal structures over reproductive success (or its components). I elaborate on how this idea for the causal distinction made below allows one to make sense of the evolution of novelties by niche construction as evolution by selection for them.

The heterogeneous distribution of types to environments can arise by chance, but it can also arise, as we have seen, because of a probabilistic association between types and their contexts. In order to save the idea that selection requires a common environment, underlying Brandon's (1990) reasoning about such cases seems to be the following considerations. Suppose there is a positive probabilistic association between two causes of reproductive success (or its components). One of the causes, N_i, negatively causes F_i whereas the other cause, N_{ij}, positively causes F_i so that type 1 in a neighborhood with mostly type 1s has a higher expected number of offspring than does a type 0 in a neighborhood with mostly type 0s. The next two-step procedure picks out which of the variables evolves by selection and which evolves as a by-product of selection. First, begin with one of the variables, say, N_i, and ask which type is favored

given its direct influence on number of offspring—ignoring the indirect association between the considered variable and the number of offspring. In this case, type 0 is favored. Then take into account the indirect association and ask which type is favored. In this case, the type 1 is now favored because, though type 1 decreases its own expected number of offspring by 3, for every type 1 in its neighborhood, it gets an additional 67 expected offspring. Second, begin with the other variable, N_{ij} and ask which type is favored given its direct influence. In this case, type 1 is favored. Then take into account the indirect association and ask which type is favored. In this case, type 1 is still favored. Because the ordinal relation among the expected number of offspring changes in the first step of the procedure and does not change in the second step of the procedure, there is a selection process on N_{ij} for type 1. Type 1 evolves with respect to N_i only because that is a by-product of selection on N_{ij}.

Whether or not the reconstruction of Brandon's argument here can be attributed to him, the conclusion is mistaken because the probabilistic association between individual type N_i and neighboring type N_{ij} can arise in two ways, ignoring common cause cases. (i) N_i (for $i = 1-4$) could directly cause N'_{i1} (or L) during dispersal, or either (ii-a) N_i could directly cause N_{ij} (for $j = 1-4$) during selection or (ii-b) could indirectly cause N'_{i1} through F_i as in side a of our population. Case (i) is discussed extensively in Glymour (2011) as a representation of developmental or phenotypic plasticity. The new type at the formerly vacant site is (positively) caused by the types of its four von

Neumann neighbors. Cases (ii-a, ii-b) involve niche construction. In (ii-a) before selection, types construct their von Neumann neighborhood directly whereas in (ii-b), types construct their neighborhood piecemeal and indirectly through dispersal.

Given the causal distinction between (i) and (ii-a, ii-b), the non-homologous type 1 is not a free-rider on selection for neighborhood type. For that to be true, we would need to replace the causal structure (ii-b) in side a of our population by causal structure (i). For example, we could change dispersal from local to global and allow developmental plasticity to influence L upon the formerly vacant site's occupancy. Then in a subsequent round of local selection that individual's type is represented by N_i. Because of developmental plasticity, there is a positive association, we have supposed, between N_i and N_{ij} (for $j = 1-4$). Given this and our causal parameters, non-homologous type 1 should become prevalent because of the large positive association between N_i and F_i, not as the result of the causal influence of N_i on F_i but, as a result of the positive influence of N_{ij} (for $j = 1-4$) on both N_i (during dispersal) and F_i (during selection). Yet, this is not what we have. Rather we have N_i causing N'_{i1} through F_i in which case during subsequent rounds of local selection, types cluster together. Given this, the non-homologous type 1 becomes prevalent as we have seen. This is because (during dispersal) N_i has a positive causal influence on N_{i1}. So, in a subsequent round of local selection that new individual's type is represented by N_{ij} (for $j = 2-4$), and (during selection) that neighbor's type has a positive causal influence on F_i. Hence, non-homologous type 1

evolves by selection for it and not as a by-product of selection for neighborhoods abundant in type 1s.

5. Implications

A number of implications arise if the argument above is correct. Obviously the intuition that a unique common environment is required for selection to drive adaptations or novelties to fixation is mistaken. Given that novelties arising by niche construction are selected for when types are heterogeneously distributed into contexts, there is no sense in which the adaptation or novelty evolves within a unique common environment. Less obviously, our concept of drift as the biased or heterogeneous distribution of types into contexts is mistaken. To suggest a way in which our concept of drift can be reconceptualized, I rely on a framework constructed by Glymour (2011).

Glymour (2011) argues for the importance of explicitly representing the causes of reproductive success in diagnosing when a character or trait is an adaptation to specific conditions and for accounting for phenotypic plasticity and local coadaptation. This strategy is understood as having two essential features:

> First, the resulting models employ variables that quantitatively characterize the… features that causally influence survival and reproductive success. Second, the causal influence of these variables is explicitly represented (Glymour 2011, 454).

Glymour (2011, 460-461) distinguishes between two ways of thinking about the causes of reproductive success—narrowly and widely. Causes of reproductive success are modeled in the narrow sense when represented as a frequency distribution over the values for the variables causing reproductive success. Causes of reproductive success are modeled in

the wide sense when represented as a probability distribution over the values for the variables causing reproductive success. With this framework, we can now initially answer the question of what evolution by drift consists in.

Drift is often described with analogies to coin flips. The frequency of heads in a run of independent flips cannot be deduced from the probability of heads. Sampling error creeps in given a finite number of flips; the frequency of heads deviates from the probability, and this is drift in the most general sense (see, e.g., Pfeifer 2005; Sober 1984). We can also say, I think quite reasonably, that sample error or drift matters when what is expected to occur conditional on the probabilities does not occur. For the sake of some concreteness, let us think in terms of our running example. Assume that character type causes neighboring types if and only if a hypothetical change in an individual's character type would change the probability distribution over the values for the neighboring types of that individual (as in niche construction). Further, assume that whether there is a causal connection between character type and neighboring types or not, this can be represented in terms of the conditional probability distribution $\Pr(N_{ij} \mid N_i)$.

This is our wide conception of the causes of (components of) reproductive success, relative to our case. The causal connections between character types and number of offspring, and neighboring types and number of offspring can also be represented so that individual conditional expected number of offspring can be estimated. Finally, for our particular case, let the narrow conception of the causes of (components of) reproductive success be $\mathrm{Freq}(N_{ij} \mid N_i)$. With these somewhat concrete ideas in mind, I say the

following more generally about drift. We have evolution by drift when the causes of reproductive success—narrowly conceived—deviate from the wide conception of the causes of reproductive success to such an extent that what is expected to evolve given the causes of reproductive success under the narrow conception is not what is expected to evolve given the causes of reproductive success on the wide conception.

Given this conception of drift and the conception of novelties evolved by niche construction argued for in the last section, a final implication of my argument is that we can diagnose when a character goes to fixation by drift or by selection for it only if we explicitly model the causes of reproductive success as well as the causal connections among those causes of reproductive success in both the wide and narrow sense. The evolution of a character or trait cannot be understood as the result of drift or selection unless causes of reproductive success are modeled both as frequencies of values for variables that cause reproductive success and as probabilities of values for variables that cause reproductive success. This is quite interesting because it is often supposed that whether drift or selection dominates in the evolution of a population can be read off from the product of a selection coefficient (a function of fitness parameters) and effective population size. Many will admit of course that analysis of this product is but a rule of thumb, and I do not doubt that most of the time it is a useful one at that. However, if the arguments above are correct, the rule of thumb's reliability fundamentally rests on how the narrowly modeled causes of reproductive success relate to and deviate from the causes of reproductive success modeled on the wide conception.

6. Summary

Many proponents of the Extended Synthesis argue that niche construction is a powerful driver of the evolution of novelties. Yet the results of their work have been under-appreciated in part, I suspect, because the traditional frameworks of the Modern Synthesis have trouble making sense of characters selected for in heterogeneous environments. I argued that the evolution of novelties by niche construction could be construed as the selection for the novelties by giving up the notion that selection requires a common environment. If we choose to go this route, then implications concerning our conception of drift and the practice of sorting episodes of evolution by selection and episodes of evolution by drift almost immediately follow.

CAUSATION, GROUP SELECTION, AND PLURALISM

1. Introduction

Damuth and Heisler (1988) take group selection to occur if and only if individual reproductive success varies and group membership causally influences individual reproductive success. Differently put, Goodnight et al. (1992) state that multilevel selection occurs if and only if variation in the fitness of individuals is due to both the properties of individuals and the properties of groups of which they are members (see also Sober 1984). Heisler and Damuth (1987) endorse a way to estimate the strength of group selection that involves the following two steps. First, define a demographic variable D_i as some moment of the distribution of an individual trait or character variable Z_i over a collection of individual to which individual i is mapped. Second, regress reproductive success on D_i controlling for Z_i. If the partial regression coefficient on D_i is statistically significantly different from zero, we can infer that there is group selection. This two step strategy is common place in evolutionary biology studies of group selection. Group selection is determined by the relevance of demographic variables to the reproductive success of individuals rather than by the need to attribute fitnesses to groups. This is the causal theory of group selection, and it is in stark contrast to the traditional non-causal notion in which what is required for selection at any level of biological hierarchy is variation in fitness at that level of biological hierarchy.

Recently, Godfrey-Smith (2008) has argued against the causal theory of group selection. In this chapter I defend the causal theory of group selection against Godfrey-

Smith's challenges. The order of discussion is as follows. In section two, I define the causal theory of group selection more carefully and give an example of a population undergoing (a kind of) group selection. In section 3, I characterize Godfrey-Smith's argument against the causal theory of group selection. His conclusion prohibits modeling the causal influence of the kind of demographic variables in our example population. The premise on which his argument relies and on which the ban on modeling certain variables is a consequence is the following: Because group selection requires variation in group fitness, we do not have group selection unless groups form non-arbitrary equivalence classes. In section 4, I argue that failure to model the causal influence that demographic variables have on individual reproductive success in our example results in predictive error. Given this, we have two options. First, we might make no claim about which theory of group selection is correct but reject Godfrey-Smith's criterion of non-arbitrary equivalence classes. Second, we might reject Godfrey-Smith's criterion and with it the claim on which it rests, namely, that group selection requires variation in group fitness. I take the first route here, as I have no argument that the non-causal theory of group selection is false. Yet I do explore the implications (for the thesis of model pluralism) of rejecting the non-causal theory of group selection (i.e., of assuming that the causal theory of group selection is correct) in section 5.

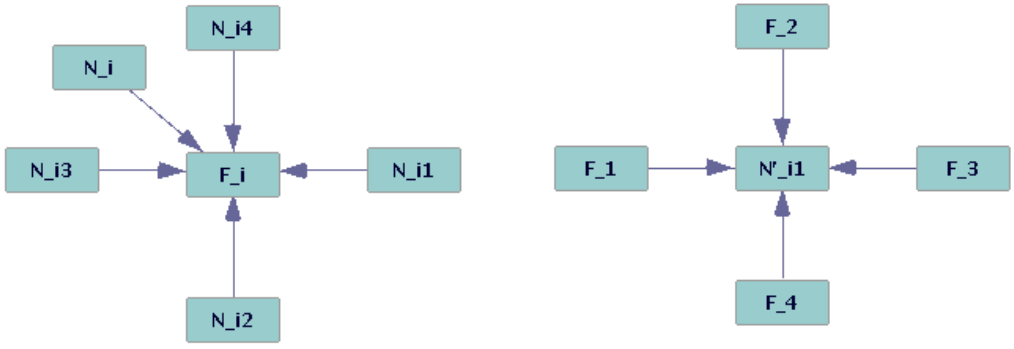
2. Causal Theory of Group Selection

Here I first characterize a version of the causal theory of group selection. The following ideas might not exactly capture what Damuth and Heisler (1988) and Goodnight et al. (1992) had in mind, but part of the motivation of this chapter is to see how far their novel

conceptions of group selection can be pushed and remain fruitful. In particular, I wish to expand their ideas so that demographic variables might include neighborhood variables that measure the type N_{ij} of a, say, von Neumann neighbor j of an individual i. Let us say the following in general about group selection:

> Group selection occurs if and only if there exists some value for a variable measuring demographic conditions so that
>
> (GS1) Groups or neighborhoods vary with respect to those demographic conditions and
>
> (GS2) Groups or neighborhoods with one kind of condition θ_i are selected for because the demographic variable causes reproductive success of individual units i and there is some set of background conditions and some alternative demographic condition θ_i' so that, under those background conditions, changing one's demographic conditions from the alternative θ_i' to θ_i would increase one's fitness.

Given the definition of group selection on offer here, let us now examine a generative model that exhibits group selection on variable N_{ij} with values n_{ij} taking the place of θ_i in the definition. Suppose we have an asexual population with two types—type 0 and type 1. The population exists on a landscape and each individual lives on a unique site or patch of the landscape. Generations are overlapping. Individuals are randomly and asynchronously chosen from the population and updated. Specifically, each



4.1. (A) Selection: The Causal Structure for Determining Number of Offspring. (B) Dispersal: The Causal Structure for Determining the Type at the Focal Site.

generation involves the following. A unit is chosen and removed. Then each unit within the dispersal distance of the newly vacated site reproduces with perfect heritability and without transmission bias, according to a function described below. From the pool of all offspring produced by units within the dispersal range, one is chosen at random to become the new occupant of the site, and the remaining offspring die. This is repeated with sites being chosen for updating the unit residing on it with replacement. Importantly, the offspring can disperse only to the von Neumann neighborhood of the parents.

More precisely, the vacant location has four von Neumann neighbors—types denoted by $N_1 - N_4$. Each of these types in turn has four von Neumann neighbors including the former occupant, numbered 1-4 clockwise starting from the former occupant. That is, type denoted by N_{ij} for $\{i, j\} = \{1-4\}$ so that, for example, N_{13} is the

type of the third neighbor of the first neighbor of the focal vacant spot. Each von Neumann neighbor then generates a number of offspring and the type at the focal vacant spot is updated by drawing at random from the pool of all the offspring created by the von Neumann neighbors characterized by N_i. The number of offspring F_i for neighbor i of a newly vacated site is determined by the following:

$$F_i = c + \alpha_1 N_{i1} + \alpha_2 N_{i2} + \alpha_3 N_{i3} + \alpha_4 N_{i4} + \alpha_5 N_i + \alpha_6 N_{i1} N_i + \alpha_7 N_{i2} N_i + \alpha_8 N_{i3} N_i + \alpha_9 N_{i4} N_i \quad (4.1)$$

a constant c, a separable contribution from the former type of the vacated site α_1, another separable contribution from each of the other von Neumann neighbor types α_{2-4}, a final separable contribution from the type of the unit i itself α_5, and a non-separable contribution from the type of the unit i itself and each of its von Neumann neighbors α_{6-9}. This is represented graphically in figure 4.1A. We will simplify the equation by assuming $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ and $\alpha_6 = \alpha_7 = \alpha_8 = \alpha_9$. Note that we are using number of offspring F_i rather then number of surviving offspring or reproductive success. However, as the new individual is chosen at random from the pool of offspring generated by all of the von Neumann neighbors of the vacated site, I contend that F_i for a generation is a reasonable proxy for reproductive success and therefore fitness. The higher the number of offspring, the better the reproductive success. In addition, the fact that in the model the next occupant is chosen by pooling the offspring of all the von Neumann neighbors entails both that F_i causes N'_{i1} after selection (i.e., during dispersal) and that there is only

one (new) occupant. In particular, the probability that N_{i1} becomes a type 1 individual is determined by:

$$\Pr(N'_{i1} = 1) = \frac{\sum_i F_i N_i}{\sum_i F_i} \qquad (4.2)$$

The probability that N_{i1} becomes a type 0 individual is given by $1 - \Pr(N'_{i1} = 1)$. The causal dependency of the new type on number of offspring is represented graphically in figure 4.1B. Given equation (4.1) and figure 4.1A, the causal theory of group selection regards this population as experiencing both individual level and group level selection. There are neighborhood effects on individual reproductive success through their effect on the number of offspring. The part of the organisms's life-cycle involving offspring dispersal ensures the possibility that the population evolves by both individual selection and group selection.

3. Godfrey-Smith's Argument

Godfrey-Smith (2008) argues, however, that it is not always true that group selection occurs if demographic variables influence the reproductive success of individuals. Indeed, his argument explicitly prohibits modeling the causal influence of the kind of demographic variable N_{ij} is in our sample population introduced in the previous section.

He arrives at this conclusion by arguing for the following criterion: Group selection occurs only if collectives of individuals can be partitioned into non-arbitrary equivalence classes. This premise, in turn, is defended by a widespread view that group selection requires (co)variance in the fitnesses of groups. If Godfrey-Smith is correct, then the

causal theory of group selection cannot be correct. In particular, it would be a mistake to model the causal influence of demographic variables like N_{ij} on individual reproductive success and call this group selection.

Demographic structure in the group selection literature is typically represented as consisting in (at least approximately) non-overlapping chunks or collectives of units wherein frequency-dependent selection occurs and between which the chunks themselves might be thought to compete (see also Okasha 2006; Sober and Wilson 1998; Wilson 1975). This assumption of non-overlapping chunks entails a fact of the matter concerning the contribution each individual makes to group level reproductive success—i.e., average individual reproductive success for the group. Indeed, such demographic structures are implicitly assumed by any Pricean covariance analysis of evolutionary change (Price 1970, 1972). Yet, our population fails on this score. I might be a part of your neighborhood, but my neighbors include individuals not in your neighborhood and so on. Therefore, in our case, I belong to more than one chunk and, as there is overlapping membership, it is not at all straightforward how my reproductive success should count toward the reproductive successes of the different neighborhoods to which I belong. Given this, nothing exists above the individual to which one can assign fitnesses such that those higher level entities can be understood as competing in the same sense the individuals comprising them do.

Godfrey-Smith (2008) argues in the following manner. He argues that multilevel selection requires that higher-level chunks or collectives of units consist in non-arbitrary

or biologically salient equivalence classes. He describes a typical instance of intrademic group selection. Individuals live in social groups that dissolve and reform each generation. That is, "groups form, selection occurs, and then the groups dissolve, creating a single 'pool' of new individuals from which a new generation of groups is formed" (2008, 27). Importantly, for Godfrey-Smith, when groups form, each individual is a member of one and only one group. Hence, there is a fact of the matter concerning what contribution each individual makes to the reproductive success of the group. Because of this, an equivalence relation—e.g., for all i, j individual j influences the reproductive success of individual i—collects individuals into equivalence classes (2008, 29-31). Consequently, groups can be assigned fitnesses and properly be seen in competition with one another, insofar as they really are.

Following this, Godfrey-Smith presents a case where a population is "arrayed on a two-dimensional lattice" (2008, 31). In this case, we cannot treat each individual's reproductive success "as a contribution to the output of a single collective entity that competes with others in a higher-level selection process" (2008, 33). A population structure with the overlapping chunks one gets from a lattice just "is not a kind that yields nonarbitrary higher-level units that can reasonably be said to compete, and that can be plausible bearers of fitness" (2008, 34). This is because there is no possible non-arbitrary equivalence relation that collects individuals into equivalence classes. From this, Godfrey-Smith concludes that there is "more than one level of selection" only if equivalence classes can be "recognized (at least approximately) in the population

structure" (2008, 36). In other words, we do not have selection occurring at multiple levels unless there exists a non-arbitrary partitioning of individuals into equivalence classes. Because there is no such partitioning for our population introduced in the last section, there is no genuine multilevel selection. Therefore, according to Godfrey-Smith, cases like ours admit only of lower-level representations—representations on which fitness (or its components) are attributed to individuals—even though demographic variables cause individual reproductive success (or its components).

Variables like N_{ij} in our generating model above violate Godfrey-Smith's criterion, and this idea, together with the idea that group selection requires (co)variance in group fitness, consequently, imposes constraints on legitimate representational models of underlying generative truths. To see this, note that for Godfrey-Smith, if we are to represent demographic conditions in our models, the variables representing those conditions must be so that for all i, j, k

(PGS1) if individual j influences the reproductive success (or components thereof) of individual i, then individual i influences the reproductive success of individual j and

(PGS2) if individual j influences the reproductive success of individual i and individual k influences the reproductive success of individual j, then individual k influences the reproductive success of individual i.

The second conditional is violated in our case. Why should we care? When variables like our N_{ij} are used anyway by representational models of selection, it is possible to diagnose higher-level selection even when groups do not compete with one another. That is, we can divide our population into non-overlapping neighborhoods and assign them fitnesses (cf. Glymour and French unpublished). Yet Godfrey-Smith explicitly calls any such divisions arbitrary and discounts such models as violations of his criterion and hence views such models as failing to meet the requirement of group selection that group's (co)vary in their fitnesses. Consequently, we must not use variables like N_{ij} for fear of countenancing group selection when there is no genuine competition among groups. If Godfrey-Smith is correct, the causal theory of group selection must be rejected and we must model our population without recourse to variables like our N_{ij}. However, if there is good reason to reject Godfrey-Smith's criterion, then the causal theory of group selection goes unharmed by Godfrey-Smiths' argument.

4. For the Causal Theory of Group Selection

4.1. Predictions Ignoring Local Demographic Conditions

It is commonly thought to be important to use multilevel models of selection to represent a population when individuals exhibit both within and between group (co)variance in reproductive success. Though, as many have noted (Kerr and Godfrey-Smith 2002; Okasha 2006; Sober and Wilson 1998), it is possible to model multilevel selection with a purely individual-level model, it is typically taken as a mistake to do so. This is because failing to represent the correct causal structure fails to meet certain predictive or

explanatory aims. In these cases the causal structure is misrepresented because the only causal claims implicit in individual-level-only evolutionary models is that type causally influences reproductive success. When local demographic conditions also cause individual reproductive success, these models are mistaken. If the failure to represent the causal influence of local demographic conditions leads to predictive or explanatory error, then we must represent such casual influences. In fact, in our population, failure to represent the causal influence of N_{ij} on individual reproductive success (or components thereof) does lead to predictive error.

To see this, let us give our population a Pricean individual-level-only covariance representation (Price 1970). We can do this by averaging over or marginalizing out particular neighbor effects. In particular, we will do this with respect to the specific neighborhood of the (former) type of the vacant spot as well as the neighborhood of the types competing for the vacant spot. Consequently, let $B(N_i = 1) = B(N_{ij} = 1)$ stand for the frequency of type 1 in the entire population and let $B(N_i = 0) = B(N_{ij} = 0)$ be the frequency of type 0 in the entire population. I will say that a type 1 competing against a single type 1 gets $F_i(1,1) = 0.25c + \alpha_1 + 0.25\alpha_5 + \alpha_6$ number of offspring and this number will be multiplied by the frequency of type 1s in the entire population, thereby averaging over the particular neighborhood of an individual occupies. Similarly, say that a type 1 against a single type 0 gets $F_i(1,0) = 0.25c + 0.25\alpha_5$ number of offspring, a type 0 against a single type 1 gets $F_i(0,1) = 0.25c + \alpha_1$ number of offspring, and a type 0 against a single

type 0 gets $F_i(0,0) = 0.25c$ number of offspring. The parameters c and α_5 are divided by four (the values of the parameters below are chosen so that this makes sense) to get the number of offspring from one round of selection (one generation) with one opponent. Also because $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ and $\alpha_6 = \alpha_7 = \alpha_8 = \alpha_9$, we just represent one of those parameters just in case one is up against a type 1 in one's neighborhood. Then a version of the Price equation for our system is the following:

$$\begin{aligned} \bar{F}\Delta\bar{N}_i &= \mathrm{COV}(F_i, N_i) \\ &= \sum_{n_i \in N_i} N_i \left(B(N_i) \sum_{n_{ij} \in N_{ij}} \left(F_i(N_i, N_{ij}) B(N_{ij}) - \bar{F} \right) \right) \\ &= B(1)\left(F_i(1,1)B(1) + F_i(1,0)B(0) - \bar{F} \right) \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} \bar{F} = {} & B(1)\left(F_i(1,1)B(1) + F_i(1,0)B(0) \right) \\ & + B(0)\left(F_i(1,0)B(1) + F_i(0,0)B(0) \right) \end{aligned} \tag{4.4}$$

is the mean number of offspring produced by the population in a generation after a space has become vacant. The evolutionary equilibria given by equation (4.3) are equivalent to those given by the replicator equation (Page and Nowak 2002). Here is one version of the replicator equation:

$$\begin{aligned} \Delta B &= B(1)B(0)\left(F_i(1,1)B(1) + F_i(1,0)B(0) - F_i(0,1)B(1) - F_i(0,0)B(0) \right) \\ &= B(1)B(0)\left((F_i(1,1) - F_i(0,1))B(1) - (F_i(0,0) - F_i(1,0))B(0) \right) \end{aligned} \tag{4.5}$$

Setting replicator equation (4.5) to zero and solving for the frequency of type 1 gives one non-trivial equilibrium point that might differ from 0 and 1:

$$\hat{B} = \frac{(F_i(0,0) - F_i(1,0))}{(F_i(0,0) - F_i(1,0)) + (F_i(1,1) - F_i(0,1))} \tag{4.6}$$

This occurs if and only if $\alpha_6 \neq 0$. Because this parameter (and similarly α_{7-9}) is nonzero, type and neighboring types interact to cause the number of offspring. Here we will explore four different causal parameter setting with interactive causation.

First, we have $c = 104$, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 25$, $\alpha_5 = 100$, and $\alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = -50$, and, second, we have the same except $\alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = -35$. In these two cases, we have symmetric and asymmetric altruistic competition, respectively. No matter what type you are, being surrounded by the opposite type confers a greater benefit for the number of offspring you produce (cf. Lanchier 2015; Sober and Wilson 1998). The second case is asymmetric altruistic competition, however, because the advantage produced from changing your type to the opposite type of your neighbor is not the same across type changes. In this second scenario, I will say that, though both types are altruistic in the relevant sense, type 1 is less altruistic than type 0.

For any "altruistic" causal influence over (the components of) reproductive success, each type has a positive probability of surviving and, so, they are said to "coexist" (Nowak 2006). Depending on the particular strengths of the causal influence of N_i and N_{ij} on F_i, there is a stable equilibrium point given by equation (4.6) between 0

and 1. In symmetric altruistic competition we have a stable equilibrium at 0.50. In asymmetric altruistic competition we have a stable equilibrium point at $\hat{B} = \frac{26-51}{(26-51)+(41-51)} \approx 0.71$, with the frequency of type 1 being higher because it is the least altruistic.

In the third case, we have $c = 304$, $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = -25$, $\alpha_5 = -100$, and $\alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = 50$, and, in the fourth, we have the same except now $\alpha_6 = \alpha_7 = \alpha_8 = \alpha_9 = 65$. In cases three and four, we have what I will call symmetric and asymmetric selfish competition, respectively. No matter what type you are, being surrounded by the same type confers a greater benefit for the number of offspring you produce (cf. Lanchier 2015). The fourth case is asymmetric selfish competition, however, because the advantage produced from changing your type to the same type of your neighbor is not the same across type changes. In the fourth scenario, I will say that, though both types are selfish in the relevant sense, type 1 is more selfish than type 0.

For any "selfish" causal influence over reproductive success, the system is "bistable" (Nowak 2006). Depending on the particular strengths of causal influences of N_i and N_{ij} on F_i as well as the initial frequencies of types, there is an unstable equilibrium point given by equation (4.6) between 0 and 1. In the case of symmetric selfish competition the unstable equilibrium point is at 0.50. In the case of asymmetric selfish competition, there is an unstable equilibrium point at

$\hat{B} = \frac{76-51}{(76-51)+(91-41)} \approx 0.38$. This means, that if type 1 is ever under half of the population or under 0.38 of the population, respectively, type 0 is predicted to evolve.

Except for the case of asymmetric altruistic competition, these predictions are in error of our generating model. In our generating model with selfish competition, types cluster into two "camps" until one type goes extinct, and types need not coexist under altruistic competition (see also Lanchier 2015; Neuhasuer and Pacala 1999). To get a sense of this, first examine the population on the one dimensional integer lattice $\mathbb{Z}$, so that types now have 2 rather than 4 neighbors—i.e., some of the parameters will be divided by 2. In the next sub-section, we will examine the original generating model in two dimensions, having properly trained our intuitions.

4.2. Revisiting The Generating Model in One Dimension

Suppose that all spaces to the left of site 0, including site 0, are of type 1 and the rest are of type 0. Then there is one interface (say, marked by "$\times$") at 0.50 on $0.50+\mathbb{Z}$ between clusters of types. Suppose we are dealing with either asymmetric altruistic competition or asymmetric selfish competition. If the rightmost type 1 is chosen for mortality first, then the type 1 next to the dead type 1 produces $0.5c+\alpha_1+\alpha_2+0.5\alpha_5+\alpha_6+\alpha_7$ and the type 0 next to the interface produces $0.5c+\alpha_1$. So, N_{i1} (for $i=1,2$) changes to 0 with probability:

$$\Pr(N'_{i1}=0)=\frac{0.5c+\alpha_1}{(0.5c+\alpha_1)+(0.5c+\alpha_1+\alpha_2+0.5\alpha_5+\alpha_6+\alpha_7)} \qquad (4.7)$$

where parameters c and α_5 are divided by 2 because there are 2 nearest neighbors now rather than 4. Alternatively, if the leftmost type 0 is chosen for mortality first, then this N_{i1} changes to 1 with probability:

$$\Pr(N'_{i1}=1)=\frac{0.5c+\alpha_2+0.5\alpha_5+\alpha_7}{(0.5c+\alpha_2+0.5\alpha_5+\alpha_7)+(0.5c)} \tag{4.8}$$

The dynamics depend only on the magnitude of these probabilities. If equation (4.7) is less (greater) than equation (4.8), then the interface between the types advances stochastically to the right (left) so that type 1 (type 0) goes to fixation eventually. Whether we have asymmetric altruistic competition or asymmetric selfish competition, then, the type that is the least altruistic or most selfish (in the relevant senses above) wins.

With this in mind, let us now look at scenarios of asymmetric altruistic competition or asymmetric selfish competition where types are initially homogeneously distributed across the integer lattice. In either competition type, there are multiple interfaces that behave like the one just examined (the least altruistic or most selfish type expands, creating ever larger clusters) except in two cases. First, when both neighbors of the vacant space are opposite in type of the now deceased individual, the probability that the new type is not the same as the type of the now deceased individual is equal to one. Hence, these interfaces tend to disappear. This leaves just one final interesting case. Assume the interface is between three (or more) type 0s in a row and two type 1s so that we have something like the following picture:

$$...-?-?-?-0-1-1\times 0-0-0-?-?-?-...$$

where ? can be of either type and $\times$ is the considered interface on $0.5+\mathbb{Z}$ between clusters of types. In our asymmetric cases, type 1 is either the least altruistic or most selfish whether it is engaged with type 0 in altruistic or selfish competition. Because of this, if the type 1 is chosen for mortality at the considered interface, type 1 is predicted to win, and, hence, there tends to be no evolution of the interface. If the type 0 is chosen for mortality at the considered interface, then the type 1 next to the interface produces $0.5c+\alpha_2+0.5\alpha_5+\alpha_7$,[13] and the type 0 next to the dead type 0 produces $0.5c$.[14] So either type 0 stochastically wins in the case of asymmetric selfish competition and, hence, nothing happens to the interface, or type 1 stochastically wins in the case of asymmetric altruistic competition and, hence, the interface jumps once to the right. But even in asymmetric selfish competition, it is slightly more likely that type 1 unexpectedly wins the site if the type 0 dies first (i.e., $\frac{142}{142+152}\approx 0.48$) than it is that type 0 unexpectedly wins the site if the type 1 dies first (i.e., $\frac{127}{127+142}\approx 0.47$). Given this and the fact that for all other possible interfaces the least altruistic or most selfish type stochastically prevails, types cluster and the most altruistic or least selfish type eventually goes to extinction on $\mathbb{Z}$, whether we have asymmetric altruistic competition or asymmetric selfish competition.

---

[13] This equals 93 for asymmetric altruistic competition and equals 142 for asymmetric selfish competition.

[14] This equals 53 for asymmetry altruistic competition and equals 152 for asymmetric selfish competition.

Now let us examine what happens in the cases of symmetric altruistic competition and symmetric selfish competition; i.e., where types are equally altruistic and equally selfish. Assume that the types start out in equal frequencies and that, initially, types are homogeneously distributed. Let the set of interfaces be defined as points $\times$ on $0.50 + \mathbb{Z}$ between two different types. These interfaces behave like a system of annihilating symmetric random walks. Each interface jumps one unit to the left or one unit to the right with (overall) equal probability because the individuals next to each other become identical in type. Moreover, because the probabilities favor the annihilation of two interfaces next to each other instead of one of those interfaces becoming closer to an interface greater than one unit away on $0.50 + \mathbb{Z}$, this process also promotes clustering of types. Because types start out in equal frequencies, in the short run both types find themselves in large clusters surrounded by their own type. Importantly, to stress again, this is true of either symmetric altruistic competition or symmetric selfish competition.

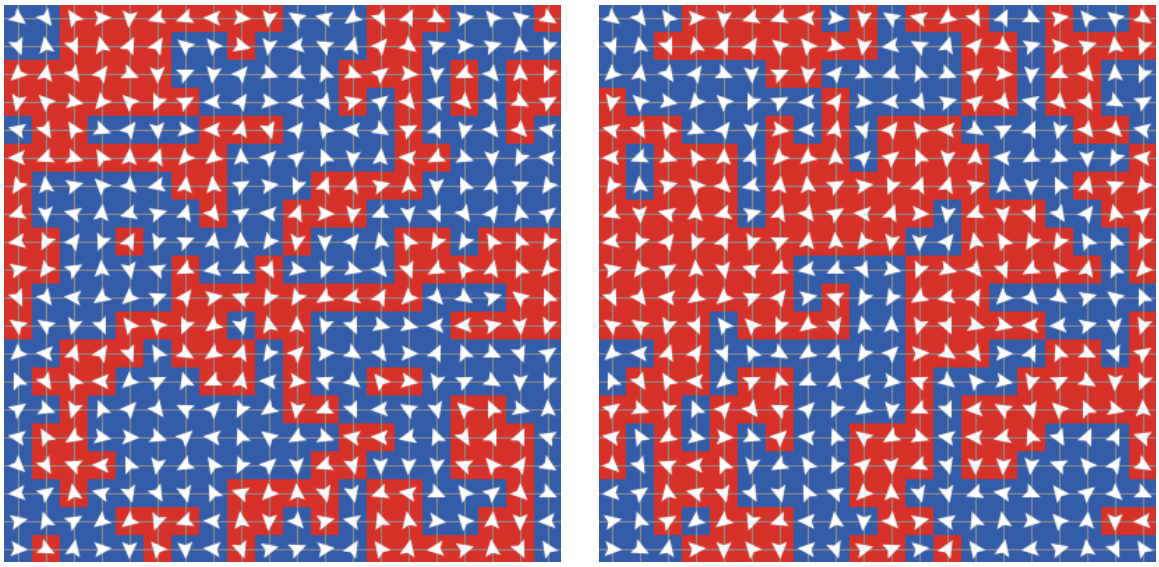
Given that symmetric random walks are recurrent in one-dimensional space, the interfaces go extinct eventually and one type fixes (though, this takes longer in the case of symmetric altruistic competition). To see this, assume for ease that all spaces to the left of site 0, including 0, are occupied by an individual of type 1 and the rest are occupied by an individual of type 0. Further, assume that—without loss of generality—the single interface at 0.50 first jumps to the right. The event that it never returns to 0.50 is the event that it reaches 2.50 before 0.50 starting from 1.50, then 4.50 before 0.50 starting from 2.50, then 8.50 before 0.50 starting from 4.50, and so on. By symmetry and the fact that types start out in equal frequencies, all jumps have probability one half. Hence, the event

that it never returns to 0.50 has probability zero. Given this recurrent property of symmetric random walks and the fact that our types start in equal frequencies, homogeneously distributed, then points along $0.50 + \mathbb{Z}$ are visited by the interfaces infinitely often and the interfaces eventually annihilate each other and one type goes to fixation. It follows from this that if one of the types starts out with greater frequency than the other, then with probability proportional to this frequency, that type will go to fixation.

To summarize the analysis up until this point, in one-dimensional space, either the least altruistic or most selfish type creates clusters of increasing size until it goes to fixation or—in the event that the types are equally altruistic and equally selfish—the set of interfaces between types behaves like a system of annihilating symmetric random walks. In the latter case, because each point on $0.50 + \mathbb{Z}$ is visited infinitely often, the set of interfaces goes extinct eventually meaning that one type goes to fixation with probability proportional to its initial frequency. Importantly, in the case where the population lives on the integer lattice, there are no non-trivial equilibria in which both types are present, whether we have altruistic or selfish competition. The only equilibria for a type are frequencies 0 and 1.

4.3. The Original Generating Model in Two Dimensions

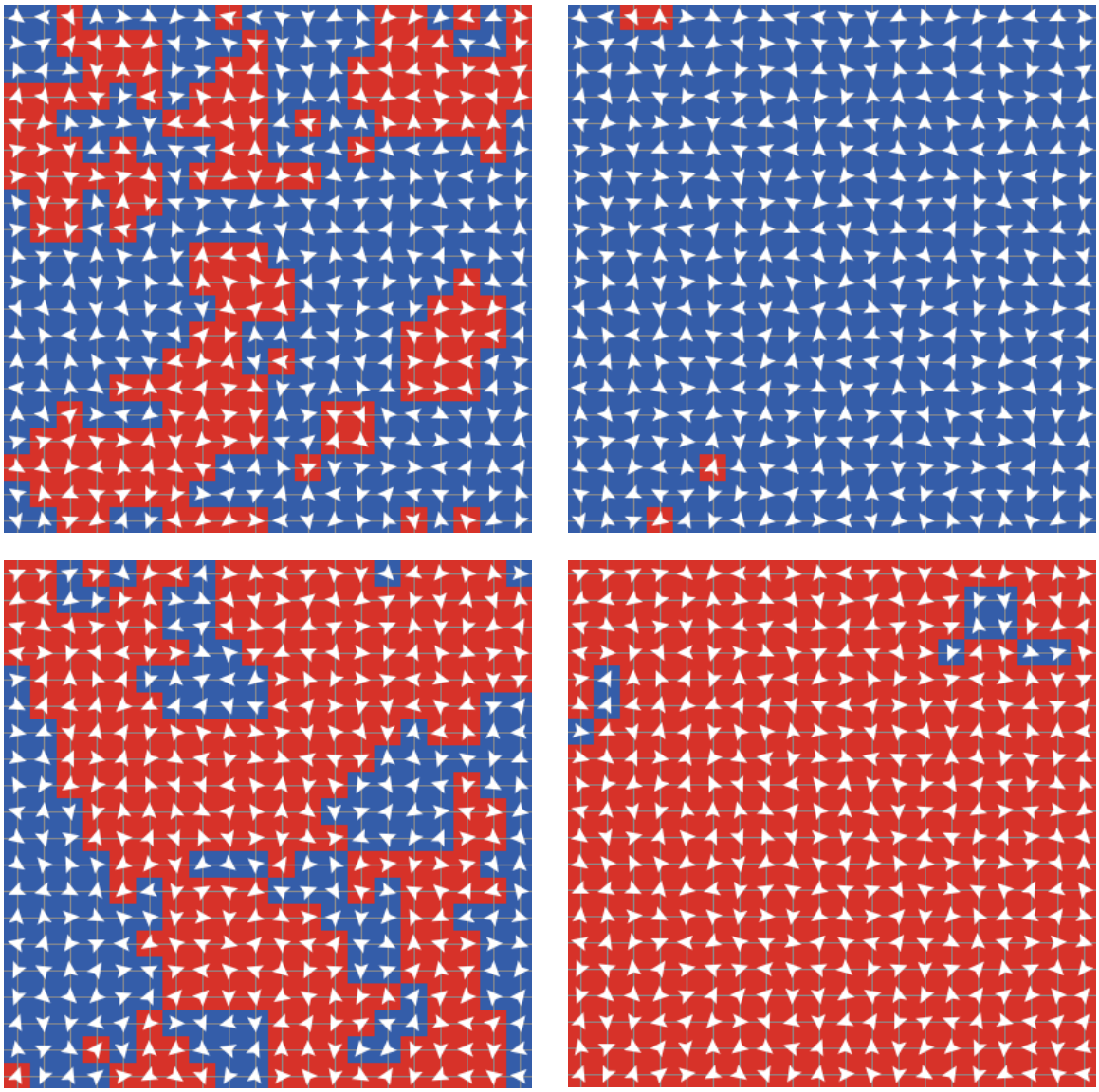
In order to explore the two dimensional case, I conducted simulations. I modeled the population on a 20×20 torus, allowed the model to run until 800,000 generations (unless fixation occurred before), and (unless otherwise stated) initiated the model so that types began in equal frequencies, homogeneously distributed. For visualization, a space with a



4.2. (A) Population 10 at Generation 20,000. (B) Population 10 at Generation 70,000.

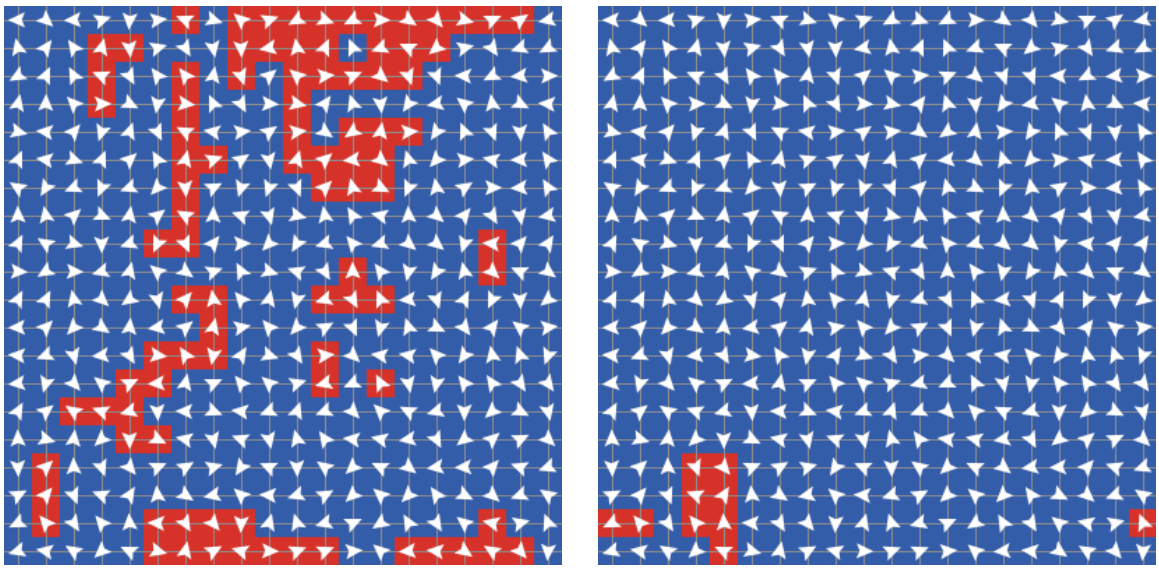
type 1 was shaded blue and a site with a type 0 was shaded red. Individuals are white arrowhead looking creatures in the figures.

I simulated the symmetric altruistic case 100 times to get an ensemble of such populations. Of the 100 populations zero of them were characterized by a type going to fixation at the end of the 800,000 generations, suggesting either coexistence or clustering. Figures 4.2A-B show population 10 at generation 20,000 and 70,000, respectively. For all populations and all generations, the world looks roughly the same as in these figures 4.2A-B. In the pictures, note that there are no large blue or red clusters but rather roughly equal numbers of blue and red. Coexistence (rather than clustering) seems to be occurring in the population. This conclusion is easier to see by juxtaposing this case with the case of symmetric selfish competition, which clearly exhibits clustering.



4.3. (A) The Top-Left Picture Depicts Population 10 at Generation 5,000. (B) The Top-Right Picture Depicts Population 10 at Generation 30,000. (C) The Bottom-Left Picture Depicts Population 25 at Generation 5,000. (D) The Bottom-Right Picture Depicts Population 25 at Generation 85,000.

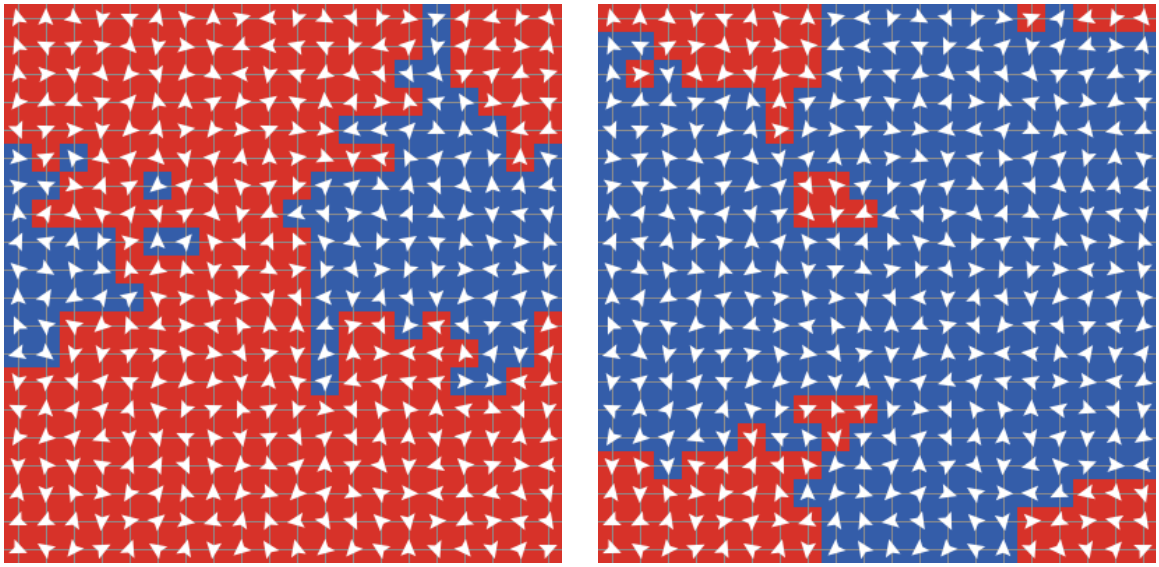
I simulated the symmetric selfish case 100 times to get an ensemble of such populations. Of the 100, roughly half of them were characterized by type 1 going to fixation before the end of 800,000 generations, and roughly half of them were characterized by type 0 going to fixation. Here we have clustering (with eventual



4.4. (A) The Left Picture Depicts Population 70 at Generation 5,000. (B) The Right Picture Depicts Population 70 at Generation 25,000.

fixation) rather than coexistence. Whichever type clusters more first wins. Figures 4.3A-D show population 10 at generation 5,000 and 30,000 and population 25 at generation 5,000 and 85,000, respectively. Unlike the previous figures, one type is clearly dominated here. Hence, in symmetric selfish competition, we still seem to have clustering with eventual fixation of one of the types, but, in symmetric altruistic competition, we now seem to have good reason for thinking coexistence is possible.

However, let us examine what occurs during asymmetric altruistic competition. Recall that under this casual parameter setting, type 1 is the least altruistic of the two altruistic types. Out of 100 populations of this type, in 100 of them, type 1 went to fixation before 800,000 generations were up. Figures 4.4A-B show population 70 at generations 5,000 and 25,000, respectively. This population is exactly like the others. The



4.5. (A) The Left Picture Depicts Population 90 at Generation 30,000. (B) The Right Picture Depicts Population 90 at Generation 45,000.

less altruistic type takes over the population quite quickly and eventually goes to fixation well within the 800,000 generation time limit.

Finally, we come to our case of asymmetric selfish competition. I simulated 100 such populations with the more selfish type 1 starting out at an initial frequency of 0.15. Of the 100 populations, type 1 fixed in roughly 65 of them by the end of the 800,000 generations. Figures 4.5A-B show population 90 at generation 30,000 and 45,000, respectively. Despite starting at such a low initial frequency, type 1 quite reliably forms larger and larger clusters until it fixes because it is the most selfish type. Hence, we get stochastic dominance of the most selfish type 1 in our asymmetric selfish competition case. Therefore, it is only in our case of symmetric altruism that we get the predictions right about our generating model when ignoring local demographic conditions.

There are, of course, other ways to model demographic conditions without modeling their causal influence on reproductive success. For example, Godfrey-Smith (2008) suggests that we might use contextualized fitness parameters where the relevant contexts are how many neighbors of a type you have in your locale. But this is also problematic as it fails to represent whether it is neighboring types or something else correlated with neighboring types that has an influence on individual reproductive success. With contextualized fitnesses, then, we might get the correct predictions of natural equilibria, but our predictions about what would happen under interventions of the system will be less generally reliable than if we were to represent the causal influence of demographic conditions. Given this and what I have shown in this section, then, I think we ought to reject Godfrey-Smith's criterion that groups must form non-arbitrary equivalence classes and with it his imposition against neighborhood variables like N_{ij}.

The causal theory of group selection survives at least this attack from Godfrey-Smith.

5. Against Model Pluralism

To be clear, I do not take myself to have offered an argument in favor of the causal theory of group selection over the non-causal traditional one. Rather I have defended the casual theory of group selection from Godfrey-Smith's attack. In particular, I have argued against his particular criterion that groups must form non-arbitrary equivalence classes by showing that the consequence of holding this view about groups (in addition to the view that group selection requires variation in group fitness), leads one to make bad recommendations about how to model certain kinds of populations where demographic

conditions cause individual reproductive success. Consequently, even if group selection does require (co)variance in group fitness, this need not arise from, what Godfrey-Smith calls, non-arbitrary equivalence classes. The causal theory of group selection, then, would just be seen as giving the necessary and sufficient conditions to generate the epiphenomenal (co)variance in group fitness that is typically used to diagnose group selection.

However, suppose that in addition to the falsity of the non-arbitrary non-equivalence class criterion, the premise underlying it—namely, that group selection requires variation in group fitness—is also false. Indeed, we might take the argument against Godfrey-Smith's criterion to further imply that group (co)variance in fitness is unnecessary for group selection. But even without an argument to that effect, if the causal theory of group selection is true, there is an interesting implication of it conjoined with my argument above that is worth exploring. That is, given what I have argued in the previous section and assuming the truth of the causal theory of group selection, we must reject model pluralism about the levels of selection (Wilson 2003).

Model pluralism is broadly the thesis that for any representational multilevel model there is a representational individual level model that makes equivalent predictions about a generative model with genuine multilevel selection and, hence, there is no fact about the generative model that is relevant for making predictions (say, about equilibria) that is represented by a multilevel model and not an individual level model. The idea, then, is that one can make predictions using either or both representational models of a class of generative models with genuine multilevel selection. Though there have been

some challenges to the thesis of model pluralism (Brandon and Nijhout 2006; but see Weinberger 2011), it remains the received view. Certain renditions of it can be found in Kerr and Godfrey-Smith (2002), Okasha (2004, 2006), and Sober and Wilson (1998, 2002).

For example, Sober and Wilson (1998, 2002) hold a thesis of predictive equivalence and model pluralism about, on the one hand, representational models that contain group fitnesses and, on the other hand, representational models that average over or marginalize out demographic conditions or neighbor effects. Sober and Wilson defend the idea that both can be explanatory, depending on the causal details of the generative scenario; yet, both can always be used interchangeably, no mater what the causal details, for (non-counterfactual) prediction.

Acceptance of such model equivalences and the ensuing model pluralism is not surprising. This is not only because many philosophers and biologists still think in terms of categorizing selection phenomena on the basis of statistical formulae, but it is also because most have implicitly assumed that the generative model approximately corresponds to intrademic group selection (Wilson 1975). Here individuals have a life cycle comprised of both selection or competition among intragenerational group members and dispersal of offspring. Selection is dependent on local frequencies of types within trait-groups (Wilson 1975, 143). Yet, the movement occurring during the dispersal stage is global. That is, types experience a period of global mixing so that, though types are spatially restricted during selection, the dispersal stage lets types transcend former spatial borders. Now, given the assumptions that groups are large and that group

assortment occurs at random, with whom an individual engages in the next round of selection depends only on the overall frequencies of types. Hence, from the point of view of evolutionary equilibria the population might as well have been demographically unstructured. The predictive equivalence thesis, on which model pluralism relies, is not surprising given the assumption that we are dealing with a deme.

However, as I have shown, the predictive equivalences are not universal and, hence, model pluralism fails. The evolutionary outcomes described in section 4 by the replicator equation differ from those predicted by a model that fully represents the causal influence of demographic conditions. The results suggest that an individual level model that leaves out neighbor effects by averaging over them will make incorrect predictions, whereas a multilevel model that explicitly represents how changing values of neighboring types influences reproductive success (or its components), will make better predictions. In particular, the individual-level-only model will predict bistability in cases of selfish competition and coexistence in cases of altruistic competition. Yet, there would seem to be a multilevel model that represents our system but never predicts bistability and will make predictions of coexistence only for some altruistic competitions. Hence, model pluralism at least about equilibria predictions must be false.

One might object, however, that the equivalence and pluralism claims of other authors have been left untouched. For example, Kerr and Godfrey-Smith (2002) defend the view that representational models assigning fitnesses to both groups and individuals and representational models assigning contextualized fitness to individuals predict identical evolutionary outcomes for a very general set of generative multilevel scenarios.

Moreover, they hold that switching back and forth between these representational models can be useful for theorizing and making predictions but that there is no feature of either model that is not fully translatable into the other and vice-versa. Hence, my argument above does not touch this rendition of model pluralism.

Recall, however, that the thesis of model pluralism is that for any representational multilevel model there is a corresponding individual level one that makes equivalent predictions about a generative model with multilevel selection, and, hence, we can use either as we please for predictive purposes. Given this, the problem with the objection is that Kerr and Godfrey-Smith's thesis fails to count as an argument toward model pluralism. For, assuming the causal theory of group selection, what Kerr and Godfrey-Smith defend is, rather, a thesis that says two distinct representations of multilevel selection—one where groups are assigned fitnesses and one where the influence of demographic conditions on individual reproductive success are represented—make equivalent predictions about a generative model with multilevel selection, and, hence, we can view the generative model in either way. But this is not model pluralism about the levels of selection, or, at least, not an interesting kind of model pluralism (see also Sober and Wilson 2002). Indeed, Okasha implicitly voices this concern with Kerr and Godfrey-Smith's paper when he says their representational models are both multilevel in that they embody "the idea that the overall change depends on selection at two levels" (Okasha 2006, 136). Hence, the objection is not sustained subsequent to replacing the statistical criterion for group selection with the explicitly causal one.

6. Summary

In this chapter I defended the causal theory of group selection against Godfrey-Smith's attacks and considered implications of my argument together with the causal theory of group selection. Godfrey-Smith's argument prohibits the use of a certain kind of variable in our representations of selection. I showed that he arrives at this claim via a criterion of non-arbitrary equivalence classes in models of selection. This criterion is motivated in large part by a traditional, non-causal conception of group selection. I argued against Godfrey-Smith's criterion by showing that it together with the view that group selection requires (co)variance in group fitness, might very well lead us to make bad representations of populations and bad predictions about equilibria and predictions about what happens under interventions. To avoid making these errors, I argued that we ought to reject Godfrey-Smith's criterion and his consequent ban on a certain kind of neighborhood variable. Even if group selection requires (co)variance in group fitness, it cannot follow from this that groups must form non-arbitrary equivalence classes. For if that is true, the implication is just the problematic prohibition on the modeling of those neighborhood variables. I then considered implications of my argument against Godfrey-Smith's objection to the causal theory of group selection. In particular, model pluralism about the levels of selection ends up looking false if we assume the causal theory of group selection is true.

A CAUSAL METHODOLOGY FOR EVOLUTIONARY BIOLOGY

1. Introduction

An important yet understudied aspect of selection is the causal relationships between phenotypic selection and the environment. We can think of this in one of two ways. First, the phenotypic trait variable and the environmental variable interact to cause reproductive success and we want to understand this interaction. Second, the parameter that codifies the strength of selection on a phenotypic trait variable—which is simultaneously the causal strength of the phenotypic variable on reproductive success—becomes a variable that can be caused by environmental variables and we want to know which ones. The discovery of these environmental causes of natural selection (Wade and Kalisz 1990) is rarely attempted (MacColl 2011). This is because of the great difficulty in gathering enough data to estimate the causal influence of environmental variables on selection in the wild as well as the fact that evolutionary biologists simply lack a causal methodology for making such inferences. Here I provide some of the first steps toward such a method.

Most estimates of phenotypic selection rely on methods such as the Price (1970) covariance approach and the Lande and Arnold (1983) regression approach. The methods yield estimates of selection differentials and selection gradients, respectively. (These are examined more closely in the following sections.) These parameters are typically used to identify whether selection is occurring in nature. Yet, sometimes they are also used to identify whether selection is fluctuating in spacetime and what might be causing selection to vary. When used in such a way, measurements of traits and reproductive success are available for different times or places. The most common procedure for identifying

whether selection is changing as a function of the environment has been to estimate selection differentials or gradients separately for each space and time (Siepielski et al. 2009; Siepielski et al. 2013).

This is problematic as it is possible to get different estimates of selection because of reasons that have nothing to do with how the causes of reproductive success interact to cause reproductive success: namely, drift (see, e.g., Chapter 2). Without actually measuring environmental variables and inferring their causal influence on selection we cannot in general tell the difference between when selection is fluctuating between times or places because of variation in the causes of selection and when drift *qua* the unexpected distribution of types into environments is simply making it appear as if selection is fluctuating. In general, verdicts on whether a trait evolves by selection or drift requires explicitly modeling the causes of reproductive success (see, e.g., Chapters 2-4), and this, in turn, will sometimes include modeling how selection on traits changes as a function of such causes of reproductive success. Hence, it is necessary to spend some time carefully constructing a causal methodology for evolutionary biology.

As I will argue in section 3 of this chapter, extending the Price and Lande-Arnold approaches to infer causes of phenotypic selection is inadvisable. Extending the Pricean approach would have to involve the intention to represent the strength of causation that a phenotypic variable has on reproductive success as an unconditional covariance or correlation between those two variables. However, because correlation is a symmetric dependence relation and causation is not, this extension cannot even get off the ground. We might instead extend the Lande-Arnold partial regression approach and take the

conditional covariance or correlation between a phenotypic variable and reproductive success, conditioning on all the other measured causes. However, this will not do because of the problem of unmeasured common causes. In some worrisome cases, regression-based extensions to infer the causes of selection will identify some variable as a direct cause of selection when in fact the identified variable has no influence on selection. That is, the Lande-Arnold approach yields false positives regarding which variables are direct causes of selection whenever there are unmeasured common causes. Hence, while both of these approaches might be useful for other aims as long as certain assumptions are met, they will not be generally helpful for inferring the causes of natural selection. In section 4 and 5 of this chapter, then, I introduce an instrumental variable (sometimes called 'the natural experiment') approach to avoid this problem of false positives. For the instrumental variable approach to work, as I first step, I make a few simplifying assumptions. I discuss these in greater detail in section 2 of this chapter, but some informal treatment might be helpful here.

The main assumptions I make are three-fold: (i) the causal relationship(s) between variables are linear with each variable being approximately normally distributed so that they can be standardized with mean zero and unit variance; (ii) Pearl's d-separation theorem (see also Chapter 1); and (iii) the interaction of environments and phenotypes on reproductive success can be represented by letting the parameter representing selection on the phenotypic variable become a variable caused by the environmental variable. I assume (i) is clear enough, though it is for sure a simplification of biological reality and might involve conceptual problems. I clarify (iii) in the next section but I do not argue for

it. Again, it is an assumption used as a first-step toward a solution to the problem of inferring causes of natural selection. Finally I merely state (ii) in Chapter 1 and in the next section, so the rest of the introduction here will be dedicated to giving some intuitions about Pearl's d-separation theorem.

Most accept the dictum that correlation between two variables does not imply causation. In fact, this is not quite correct. Correlation and some other assumptions can actually tell us quite a bit about causation and knowledge of causation can tell us all we could ever want to know about correlations among variables. The way we get there is through Pearl's (2000) d-separation theorem (see Chapter 1 and the next section of this chapter). Suppose, variables A and B directly cause C and C directly causes variables E and F, relative to these variables. Pearl's theorem implies that variables A and B are unassociated if we take into account information provided by no other variables in the set, but associated if we condition on any of their direct or indirect causes. This just means that information on C, for example, and information on A (or B) gives us an idea of the value of B (or A), but that A (or B) by itself does not provide information about the value of B (or A) because they are exogenous (uncaused), relative to the set of variables. Moreover, the theorem implies that variables A and F are associated if we take into account information provided by no other variables because A indirectly causes F through C, but not associated if we take into account information on C. That is, C "screens-off" the two other variables because it mediates the causal influence between them. Finally variables E and F are associated without taking into account any other variable's information (because they are commonly caused by C), but unassociated if we

take into account the information provided by their common cause C. With Pearl's d-separation theorem, we can also go the other way, from (conditional) association or (conditional) correlation to causal structure. For example, if we found an association between A and B, conditional on C but no association between them when provided with no other information on any other variables, we could infer that A and B have at least one common effect. If we found an unconditional association between B and F and found no association between them, conditional on C, we could infer that either $B \rightarrow C \rightarrow F, B \leftarrow C \leftarrow F$, or $B \leftarrow C \rightarrow F$ (assuming there are no unmeasured common causes between B and C and C and F). Moreover, we could make the same inference if we found an unconditional association between E and F and found no association between them, conditional on C. All of these inferences and more are licensed by Pearl's d-separation theorem. If we add more assumptions, we can make stronger inferences from correlation to causal structure and back. Some of these additional assumptions will be added a long the way.

2. Some Formal Preliminaries

The purpose of having a causal methodology for evolutionary biology is inferring the causal structure over a set of variables $\mathbf{S}$, one of which, I will suppose, is reproductive success (or aspects thereof) R. For simplicity, I will assume that R does not cause any variables, relative to the set of variables in $\mathbf{S}$. The other variables under consideration will include at least one phenotypic variable Z and at least one environmental/contextual

variable E. When I do not specify whether the variable measured on individuals is phenotypic, environmental, or otherwise, I denote it as X.

As a reminder, according to the causal modeling framework (Pearl 2000; Spirtes et al. 2000; Woodward 2003) assumed here, causation has three ingredients. Here is the first:

(C1) Causation is an asymmetric dependence relation between variables.

The relata of causal relations are variables, not values of variables, not properties, not objects, and not events. Bill depth, for example, might or might not cause reproductive success, but I will not say a particular bill depth size causes a particular (expected or realized) value of reproductive success. Here are the next two ingredients of causation:

> A variable X is a direct cause of variable R relative to a set of measured variables $\mathbf{S}$ if and only if
>
> (C2) there is a set of possible probability distributions over R generated by a function of direct causes of R, say these are $\mathbf{V}$, such that
>
> (C3) manipulating X, while holding all other variables $\mathbf{S}\setminus\{R\}$ constant, would change the probability distribution of R.

That is, we have a direct causal effect when and only when there exists some i and j and some set of (background) conditions—i.e., some set of values for $\mathbf{S}\setminus\{X,R\}$—such that changing X from value $X = x_i$ to value $X = x_j$ while holding all other variables except R constant would generate a change in the probability distribution of R. Direct causal relations can be represented in graphs as directed paths of length one (i.e., directed edges)

from the causal variable to the effect variable: $X \rightarrow R$. They can also be represented as structural equations:

$$R = f(\mathbf{V}) = f(X, U_1, U_2, ..., U_n) \quad (5.1)$$

Equation (5.1) represents the fact that the conditional expected value of reproductive success is given as a function of its measured direct causes (e.g., X) and its n unmeasured direct causes $\mathbf{U}$.

Part of the causal structure consists in causal parameters (cf. Didelez et al. 2010; Heckman 2000). Causal parameters codify the strength of direct causation of one variable on another or how one variable's expectation (for example) would change given a hypothetical change in the other variable. We can denote this specific causal parameter as $\frac{\Delta \mathrm{E}(R \mid X, U_1, U_2, ..., U_n)}{\Delta X} = \frac{\Delta W}{\Delta X}$. Causal parameters, consequently, can also represent selection. For example, the causal parameter given just now could represent the strength of direct selection on X. Causal parameters and, hence, parameters representing selection fluctuate if and only if a variable interacts with X to cause R. I say that

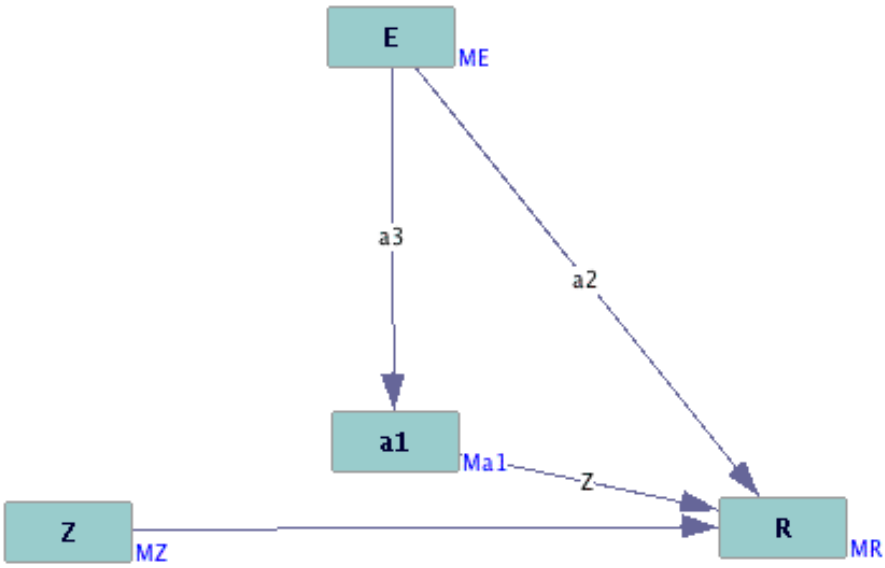
> A variable interacts with X to cause R if and only if for any pairs i and j values of X and some pair of background conditions differing only in the value of the other direct cause of R, $\frac{\Delta W}{\Delta X}$ under one background condition is not equal to $\frac{\Delta W}{\Delta X}$ under the other background condition.

Direct selection can fluctuate in three ways: $\frac{\Delta W}{\Delta X}$ can change from zero (nonzero) to nonzero (zero) (cf. Gibbs and Grant 1987; Price et al. 1984), $\frac{\Delta W}{\Delta X}$ can change sign or direction from negative (positive) to positive (negative) (Endler 1983), or $\frac{\Delta W}{\Delta X}$ can change in magnitude (Aspi et al. 2003; Kingsolver et al. 2001; Hereford et al. 2004).

In what follows, I will need to represent fluctuating selection graphically as well as with structural equations. As a first step, I assume that every causal dependency is linear and that every variable is approximately normally distributed and standardized with zero mean and unit variance. Linearity requires that the true model is (at least approximately) linear with normally distributed error term ε, with $\mathrm{E}(\varepsilon) = 0$ representing the expected value of a random draw from a function of normally distributed, standardized unmeasured variables, $f(\mathbf{U})$. So, I am assuming that equation (5.1) above can be approximated by:

$$R = \alpha X + \varepsilon_R \qquad (5.2)$$

where $\alpha = \frac{\Delta W}{\Delta X}$. The causal graph can be represented as $X \overset{\alpha}{\rightarrow} R$ with the causal parameter. I also assume Pearl's (2000) d-separation theorem (see Chapter 1). Pearl's d-separation can be stated as follows: Two arbitrary variables X_k and X_l are probabilistically dependent conditional on $\mathbf{C} \subset \mathbf{S}$ if and only if there is some path on a



5.1. An Example of Fluctuating Selection on a Phenotypic Variable. Each Variable's Mean is also Represented in the Picture and Standardized to Zero.

directed graph between X_k and X_l such that no mediator—$X_k \rightarrow C \rightarrow X_l$—or common cause—$X_k \leftarrow C \rightarrow X_l$—in the path is in $\mathbf{C}$, and every common effect or collider—$X_k \rightarrow C \leftarrow X_l$—on the path either is in $\mathbf{C}$ or has some effect variable which is in $\mathbf{C}$.

Finally, I note that in the causal modeling literature there is no agreed upon manner in which to represent interactive causation. My final assumption is that it can be represented, taking Wade and Kalisz (1990) quite literally, as a variable's causal influence on a linear causal parameter.

To see this last assumption explicitly, let us examine an example briefly. Suppose that measured, standardized variables Z and E interact to cause R. Then I represent this structurally as:

$$R = \alpha_1 Z + \alpha_2 E + \varepsilon_R \tag{5.3}$$

$$\alpha_1 = \alpha_3 E + \varepsilon_{\alpha_1} \tag{5.4}$$

The conditional expected total effect of Z on R is $\mathrm{E}(\alpha_1 \mid E) = \alpha_3 E$ and the conditional expected total effect of E on R is $\alpha_2 + \alpha_3 Z$. Graphically, this can be represented as in figure 5.1. In the figure, the reader can trace all of the directed paths from Z to R to get the conditional expected total effect of Z on R, and all of the directed paths from E to R to get the conditional expected total effect of E to R (cf. Wright 1934). Note that selection α_1 has a causal influence on R codified by $\frac{\Delta W}{\Delta \alpha_1} = Z$. This is because holding everything fixed (including $Z \neq 0$) and hypothetically changing selection would change the conditional expectation of reproductive success. Hence, that selection has a causal influence is a result of our definition of causation coupled with the assumptions made above.

3. Against Extensions of Selection Differentials and Regression Coefficients

There are many approaches to estimating what I have defined as a causal parameters representing the strength of selection (Endler 1986; Lande and Arnold 1983; Mitchell-Olds and Shaw 1987; Price 1970; Robertson 1966; Schluter 1988). I only examine two popular ones here: Selection differentials and selection gradients. Selection differentials

(Price 1970) quantify selection by the change in the mean of a phenotypic variable due to selection. This evolutionary response to selection is sometimes given as:

$$\overline{R}\Delta\overline{Z} = \mathrm{COV}(R,Z) \tag{5.5}$$

assuming perfect heritability and no transmission bias. The covariance is the selection differential. Unfortunately, the problem with the selection differential is that it simply does not qualify as a causal parameter as defined here.

Unconditional covariances or their standardized cousins—unconditional correlations CORR—quantify a symmetric relationship between variables. Yet, causation is an asymmetric dependency. In general the unconditional correlation coefficient of two variables will not be equivalent to the asymmetric causal parameter defined here. This is because of the problem of common causes. Obviously one could condition on common causes insofar as they are measured to get the conditional correlation between two variables, but now we are in the realm of selection gradients and regression-based approaches for estimating dependencies between variables. Selection gradients are consequently closer to what we want. Given this, I will call our causal parameter α a selection gradient. Nevertheless, there are problems with getting at the causes of reproductive success and selection gradients using the common un-supplemented regression-based approach.

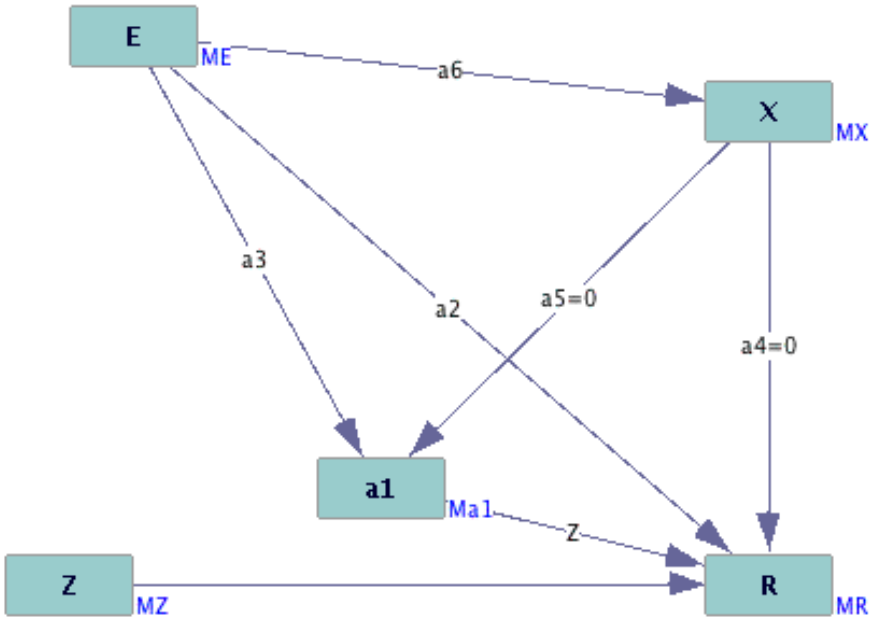
Selection gradients are often modeled as standardized partial regression coefficients (Lande and Arnold 1983), defined as:

$$\psi = \frac{\mathrm{CORR}(R,X \mid \mathbf{S} \setminus \{R,X\})}{\mathrm{VAR}(X \mid \mathbf{S} \setminus \{R,X\})} \tag{5.6}$$

In a certain sense, partial regression coefficients do not fall prey to the kind of worry noted about selection differentials. As long as one conditions on common causes, one can sometimes reliably infer direct causal dependencies between variables. Yet there is one crucial problem with estimating selection gradients with mere regression. This is the problem of unmeasured common causes.

Though textbooks and journal articles in the social sciences and epidemiology (Angrist and Pischke 2009; Antonakis et al. 2010; Cameron and Trivedi 2009; Didelez et al. 2010; Sheehan et al. 2008) spend a great deal of time worrying about just this problem, relatively little space is spent worrying about the problem of unmeasured common causes in textbooks (Burnham and Anderson 2002; Quinn and Keough 2002; Sokal and Rohlf 2011) and journal articles in evolutionary biology (but see Shipley 2000 and Walker 2014 for welcome exceptions). If we are careful to distinguish between holding variables constant or fixed and statistically conditioning on or controlling for measured variables, then the problem is clear.

Recall our model in figure 5.1 and equations (5.3) and (5.4). If we really could hold all of the unmeasured direct causes of reproductive success and the selection gradient α_1 constant, however many there are, we should expect to recover the true values of the causal parameters. Yet, for example, in regressing R on Z, we statistically condition or control for only one measured variable E. The conditioning for both equations can be written in the following manner:



5.2. Worrisome Causal Structure.

$$\mathrm{E}(R \mid Z, E) = \alpha_1 Z + \alpha_2 E + \mathrm{E}(\varepsilon_R \mid Z, E) \tag{5.7}$$

$$\mathrm{E}(\alpha_1 \mid E) = \alpha_3 E + \mathrm{E}(\varepsilon_{\alpha_1} \mid E) \tag{5.8}$$

The conditional expectation of both endogenous variables is given as a linear function of its direct causes. However, such conditioning does not, of course, guarantee that the set of unmeasured direct causes, $\mathbf{U}_R$ and $\mathbf{U}_{\alpha_1}$, of either endogenous variable is constant. The relationship generated by regression only guarantees that the set of measured variables is controlled. Unless we condition on all of the direct causes or it is the case that the

measured variables are not caused by any variables in the sets of unmeasured direct causes, it is not the case that regression will determine or identify the causal parameters reliably. Hence, extensions of regression-based procedures for inferring causes of selection gradients will not in general be reliable.

To see this, suppose the true causal model is represented in figure 5.2 and given by the following structural equations:

$$R = \alpha_1 Z + \alpha_2 E + \alpha_4 X + \varepsilon_R \tag{5.9}$$

$$\alpha_1 = \alpha_3 E + \alpha_5 X + \varepsilon_{\alpha_1} \tag{5.10}$$

$$X = \alpha_6 E + \varepsilon_X \tag{5.11}$$

where $\alpha_4, \alpha_5 = 0$. By "true" here I just mean that measuring at least E is necessary to recover the correct causal parameters with respect to the evolution of Z. For example, in equation (5.10), conditioning on E is necessary to determine the causal influence $\alpha_5 = 0$ of X on the selection gradient α_1 of interest. This is so because $\mathrm{E}(\varepsilon_{\alpha_1} \mid E, X) = 0$ only if E is conditioned on and this, in turn, is because $\alpha_1 \leftarrow E \rightarrow X$. Suppose that we have data on X but no data on E. Note that X is neither a cause of reproductive success or the selection gradient of interest. However, suppose we hypothesize that it is such a case. First, regress R on Z for ranges of values for X so that we end up with a vector of data on the selection gradient of interest, given this hypothesized cause of selection. Second, estimate the following:

$$\alpha_1 = \hat{\alpha}_5 X + \hat{\varepsilon}_{\alpha_1} \tag{5.12}$$

Note that $\hat{\varepsilon}_{\alpha_1} = \varepsilon_{\alpha_1} + \alpha_3 E$. Substituting the true causal model into the definition of the partial regression coefficient (equation (5.6)), we get:

$$\begin{aligned} \hat{\alpha}_5 &= \frac{\text{CORR}(\alpha_1, X)}{\text{VAR}(X)} = \frac{\text{CORR}(\alpha_3 E + \varepsilon_{\alpha_1}, X)}{1} \\ &= \alpha_3 \text{CORR}(E, X) \end{aligned} \tag{5.13}$$

By equation (5.11):

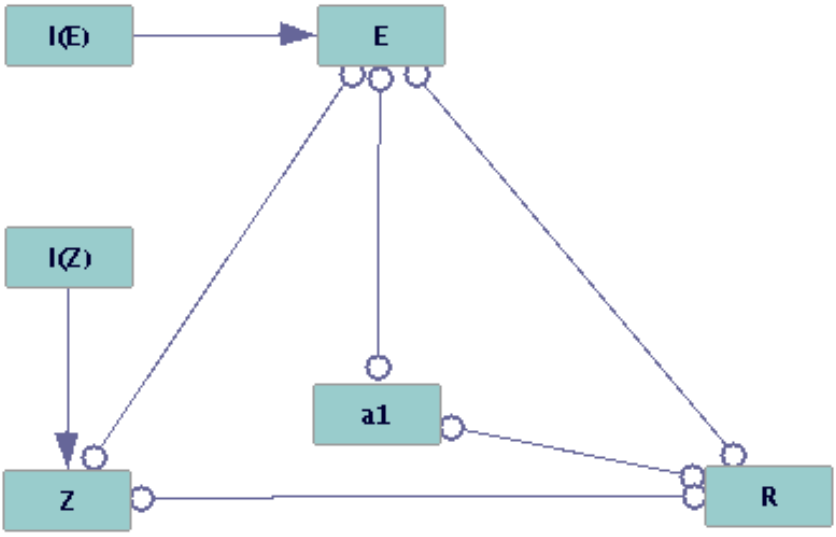
$$\hat{\alpha}_5 = \alpha_3 \alpha_6 \tag{5.14}$$

Hence, our estimate of the causal influence of X on the selection gradient is biased. It should be zero, but it is nonzero. We need a reliable way of determining the causes of selection. Standard selection differentials and partial regression coefficients will not deliver the goods in general. In the next sections, I introduce a method involving instrumental variables.

4. An Instrumental Variable Alternative

Recall, once more, the simple case of interactive causation in figure 5.1 and in equations (5.3) and (5.4). An unbiased estimate of selection requires that $\text{E}(\varepsilon_R \mid Z, E) = 0$ and an unbiased estimate of α_3 requires that $\text{E}(\varepsilon_{\alpha_1} \mid E) = 0$. If there are any unmeasured common causes, we have little reason to accept the estimates. Suppose, though, that we have two variables I_Z and I_E that satisfy the following constraints:

(I1) Relevance: I_Z causes Z and I_E causes E, relative to the set of measured variables $\mathbf{S} \cup \{\alpha_1\}$ and



5.3. Hypothetical Knowledge Graph.

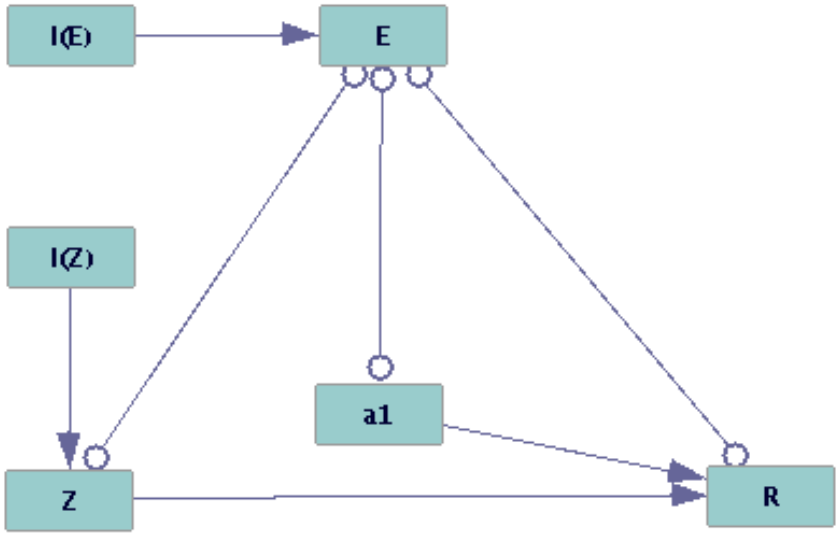
(I2) Exclusion: There are no unmeasured common causes between I_Z or I_E and any of the measured variables $\mathbf{S} \cup \{\alpha_1\}$, no measured variable in $\mathbf{S} \cup \{\alpha_1\}$ causes either of the instrumental variables, and neither of the instrumental variables cause any of the other measured variables in $\mathbf{S} \cup \{\alpha_1\}$.

The exclusion constraint (I2) implies that the instrumental variables are associated with reproductive success or the selection gradient on Z (if at all) only through the variables

they are meant to be instruments of. The relevance constraint (I1) implies that the instrumental variables are probabilistically dependent on the variables they are meant to be instruments of. I have characterized the constraints in explicit causal terms because the typical way of describing them in terms of probabilistic association is not sufficient for the work I want them to do here.

Assume we have accepted the causal knowledge that the two new variables are good instruments of the variables they are meant to instrument. Suppose we hypothesize that E is a cause of the selection gradient of interest. Other than this, we do not know the causal relations among the measured variables. Let us represent our current state of knowledge as in figure 5.3. Direct edges represent causal knowledge. No edge means we either assume or know there is no causal connection. An edge with circles on the ends represents the fact that we do not know whether there is an edge or, if there is an edge, what kind of causal connection exists. Call this last sort of edge, an "undetermined knowledge edge."

With the instrumental variables we can estimate the causal parameters between the variables in figure 5.3, overcoming the worry of unmeasured common causes. To see this, start with Z. From figure 5.3, Z might be a direct cause of R or an indirect cause of R through E. Let us see if we can rule out the indirect relation first. The total causal effect of I_Z on E is $\frac{\Delta \mathrm{E}(E \mid I_Z)}{\Delta I_Z} = \frac{\Delta \mathrm{E}(Z \mid I_Z)}{\Delta I_Z} \frac{\Delta \mathrm{E}(E \mid Z)}{\Delta Z}$. Solving for $\frac{\Delta \mathrm{E}(E \mid Z)}{\Delta Z}$ and

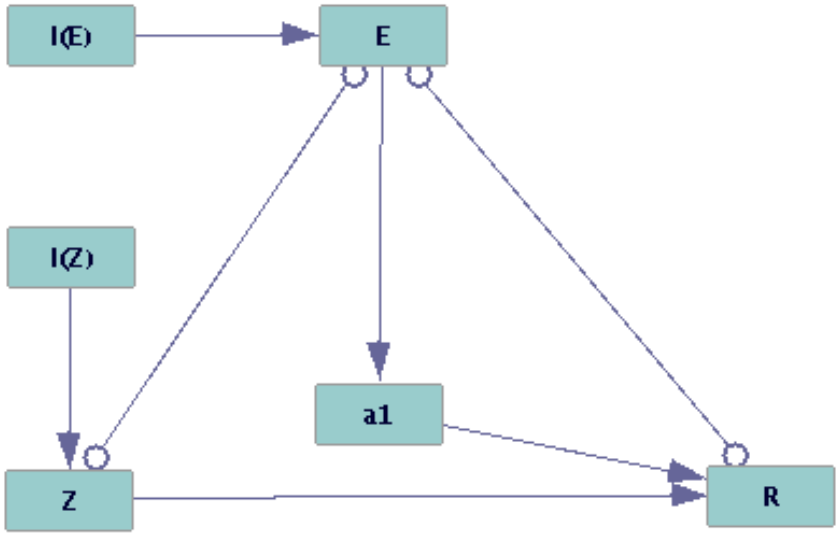


5.4. Updated Knowledge Graph.

given our assumption of linearity and the definition of a partial regression coefficient, we have:

$$\frac{\Delta \mathrm{E}(E \mid Z)}{\Delta Z} = \frac{\frac{\Delta \mathrm{E}(E \mid I_Z)}{\Delta I_Z}}{\frac{\Delta \mathrm{E}(Z \mid I_Z)}{\Delta I_Z}} = \frac{\mathrm{CORR}(E, I_Z)}{\mathrm{CORR}(Z, I_Z)} \tag{5.15}$$

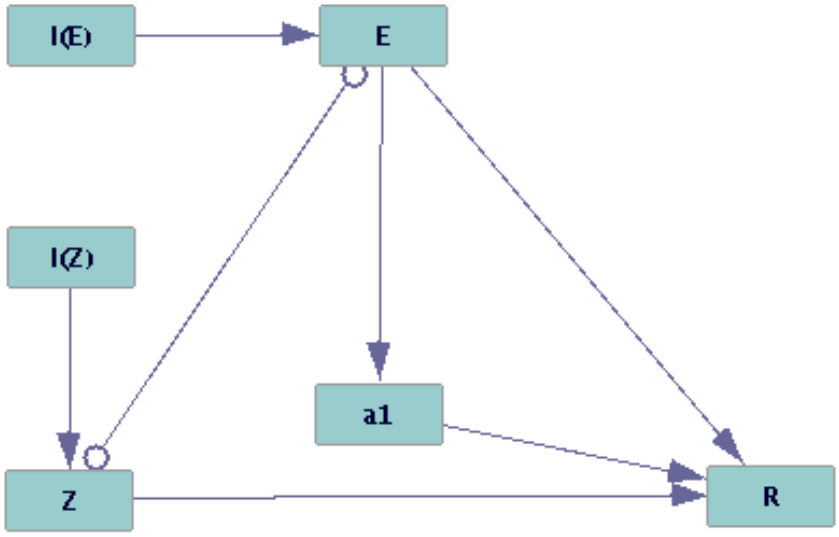
which is, of course, just the standard instrumental variable estimator. Suppose that the numerator in equation (5.15) is 0. Given our assumptions about the instrumental



5.5. Updated Knowledge Graph.

variables, this rules out the case where Z is a cause of E, though it does not rule out the case that there is a common cause between them or that E is a cause of Z. At this point, then, we do not know much more than what is represented in figure 5.3. Yet, now that we have ruled out the possibility that E is a mediator on the path from Z to R, relative to the set of measured variables, we can now estimate the strength of direct selection with:

$$\alpha_1 = \frac{\text{CORR}(R, I_Z)}{\text{CORR}(Z, I_Z)} \tag{5.16}$$



5.6. Updated Knowledge Graph.

Importantly, we can do this for different ranges of values of our hypothesized cause of the selection gradient to get a vector A of estimates of the selection gradient. Suppose that at least some of these estimates are nonzero. Then we can now update our causal knowledge to figure 5.4. In this figure, we have replaced two undetermined knowledge edges with directed edges. This is to represent that we know those direct causal influences exist. However, it is not meant to represent that we know there are no other causal connections —such as those that result from unmeasured common causes.

Next, we will want to estimate:

$$\alpha_3 = \frac{\text{CORR}(\alpha_1, I_E)}{\text{CORR}(E, I_E)} \tag{5.17}$$

Suppose that this is nonzero. So, we can replace the undetermined knowledge edge with a directed edge from the environmental cause to the selection gradient on the trait variable, as in figure 5.5. We will also want to know the direct effect of the environmental cause on reproductive success. To get this, we must first estimate $\frac{\text{CORR}(Z, I_E)}{\text{CORR}(E, I_E)}$. Suppose that this is zero. Given this, condition on $Z = 0$ and estimate:

$$\alpha_2 = \frac{\text{CORR}(R, I_E \mid Z = 0)}{\text{CORR}(E, I_E)} \tag{5.18}$$

If we do not so condition here, then we include in our estimate of the parameter, the product of $\alpha_3 Z$. Given this, we have recovered all of the causal parameters of interest (see figure 5.6).

5. A Procedure for Inferring Causes of Selection

To systematize the above example, let us introduce some useful terminology (cf. Eberhardt 2009). Define a partial ordering $\succ$ over variables X_k and X_l as follows:

> Given a set of instrumental variables $\mathbf{I}$ and a set of measured variables that might or might not include fluctuating selection gradients and of which each pair might have at least one unmeasured common cause, $X_k \succ X_l$ if and only if
>
> (PO1) $\text{CORR}(I_{X_k}, X_l) \neq 0$ and

(PO2) There is no X_m and I_{X_m} such that $\text{CORR}(I_{X_k}, X_m) \neq 0$ and

$\text{CORR}(I_{X_m}, X_l) \neq 0$.

Let a partial order graph **POG** be defined as follows:

Given **I** and a set of measured variables that might or might not include selection gradients, a **POG** is a graph where $X_k \rightarrow X_l$ if and only if $X_k \succ X_l$.

Given these two definitions, Pearl's d-separation, linearity, and the assumptions about instrumental variables, we can reliably determine the causes of selection. Here I present some pseudo-code:

1. Construct a knowledge graph over the measured variables and instrumental variables. Separate measured variables into subsets **Z** that you are interested in the evolution of and variables **E** that you hypothesize interact with the variables in the former set to cause reproductive success. At minimum, each instrumental variable should have a direct edge into the variable it is an instrument of and should not be causally connected to any other variable.
2. Begin by sorting partial orderings involving variables in **Z** and reproductive success, locally regressing R on these variables for different ranges of values of variables in **E** when considering the direct causal effect of any Z on reproductive success, where Z and any E are hypothesized to interact to cause reproductive success.

3. For each variable in $\mathbf{Z}$ within the constructed partial orderings, determine the causal effect $\beta_{I_Z Z}$ of the instrument on the trait variable it is meant to instrument.
4. For each variable in $\mathbf{Z}$ where $Z \succ R$, determine the direct causal effect of Z on R for different ranges of values of variables in $\mathbf{E}$ and generate values for direct selection A variables for these Z variables. Assign Z to the edges $\alpha \rightarrow R$.
5. Determine the direct causal effect of each variable in $\mathbf{Z}$ on each variable that is not reproductive success.
6. Next sort partial orderings involving variables in $\mathbf{E}$ and reproductive success, conditioning on $Z = 0$, for all Z where selection is either known to fluctuate (as determined in step 4) or might fluctuate (as determined in step 2), when considering the direct causal effect of any E on reproductive success.
7. For each variable in $\mathbf{E}$ determine the causal effect of the instrument on the environmental variable the instrument is meant to be an instrument of.
8. Determine the direct causal effect of each variable in $\mathbf{E}$ on each variable not in A.
9. For each variable in $\mathbf{Z}$ where $Z \succ ... \succ R$, determine the direct causal effect of Z on R for different ranges of values of variables in $\mathbf{E}$ and generate values of direct selection A variables for these Z variables. Assign Z to the edges

$\alpha \to R$. If it is the case that $Z_k \succ ... \succ Z_l \succ ... \succ R$ and selection on Z_l does fluctuate or might fluctuate, then perform this step for Z_k, setting $Z_l = 0$.

10. Determine the direct effect of variables in $\mathbf{E}$ on variables in A, generated from step 4 and step 9.

11. Orient additional direct edges and return the updated knowledge graph.

To see this better, I will apply this pseudo-code to a relatively complex example before showing why it works generally. The example involves a phenotypic variable Z_2 that directly causes another Z_1, that directly causes an environmental variable E_1, and that directly causes reproductive success R. The variable Z_1 interacts with E_1 to cause R. Another environmental variable E_2 causes E_1, and a final environmental variable E_3 has no directed causal connection with any of the measured variables. Below is the unknown structural equations among the variables with error terms left implicit:

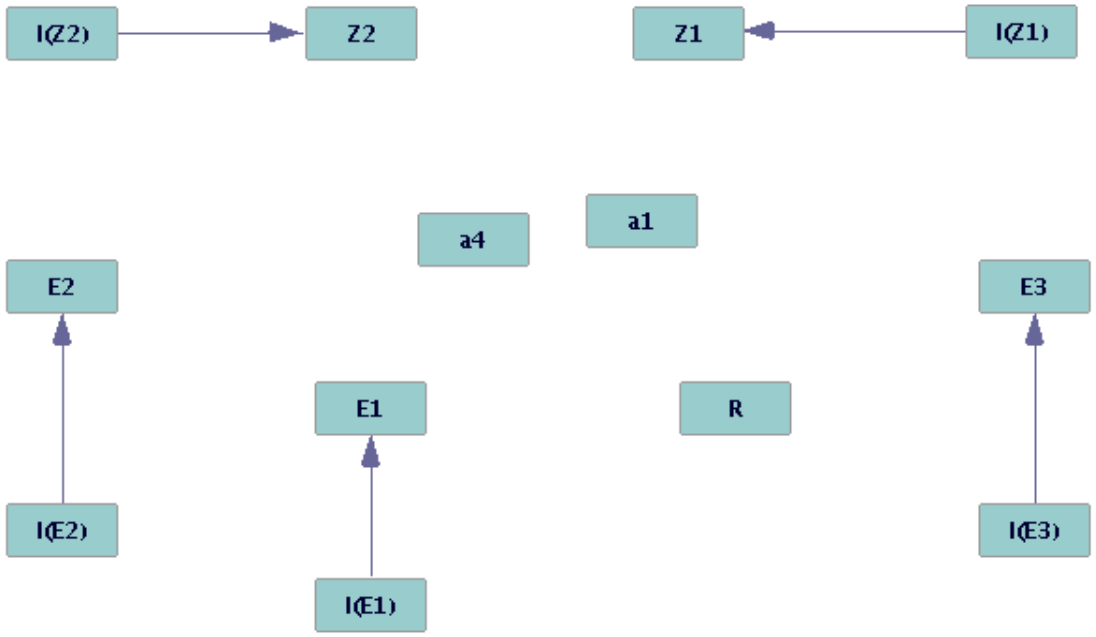
$$R = \alpha_1 Z_1 + \alpha_4 Z_2 + \alpha_7 E_1 \tag{5.19}$$

$$Z_1 = \alpha_2 Z_2 \tag{5.20}$$

$$E_1 = \alpha_3 Z_2 + \alpha_6 E_2 \tag{5.21}$$

$$\alpha_1 = \alpha_5 E_1 \tag{5.22}$$

We have a set of instrumental variables for each phenotypic and environmental variable and, in particular, we are interested in the causes of selection gradients α_1 and α_4. We hypothesize that at least one of the environmental variables interacts with the



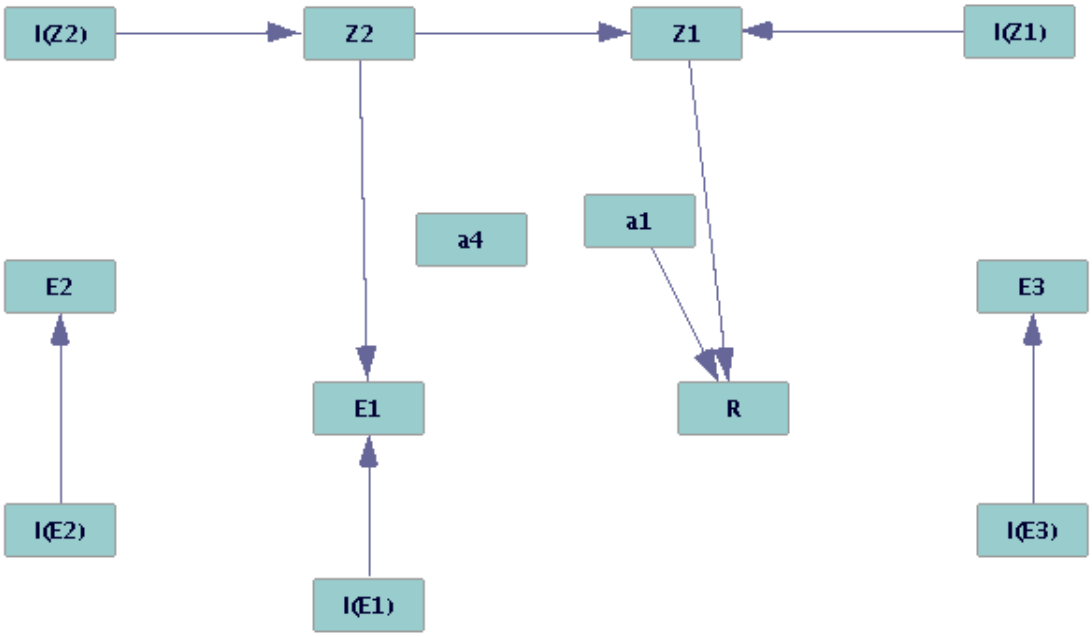
5.7. Initial Knowledge Graph with Undetermined Knowledge Edges Left Implicit to Avoid Clutter.

phenotypic variables to cause reproductive success. However, we stipulate for ease that neither of the phenotypic variables are direct causes of selection gradients.

Step 1 is to construct an initial knowledge graph as in figure 5.7. Step 2 is to sort a partial ordering involving variables Z_1, Z_2, and R. To do this we must perform the following set of standardized regressions:

$$Z_1 = \gamma_2 I_{Z_2} \quad \text{and} \quad E_1 = \gamma_3 I_{Z_2} \quad \text{and} \quad E_2 = \eta_1 I_{Z_2} \quad \text{and} \quad E_3 = \eta_2 I_{Z_2}$$
$$Z_2 = \eta_3 I_{Z_1} \quad \text{and} \quad E_1 = \eta_4 I_{Z_1} \quad \text{and} \quad E_2 = \eta_5 I_{Z_1} \quad \text{and} \quad E_3 = \eta_6 I_{Z_1}$$

Next, for some interesting set of combinations of ranges of values for each of the environmental variables, perform the following standardized regressions:



5.8. Updated Knowledge graph.

$$R = \gamma_1 I_{Z_1} \quad \text{and} \quad R = \gamma_4 I_{Z_2}$$

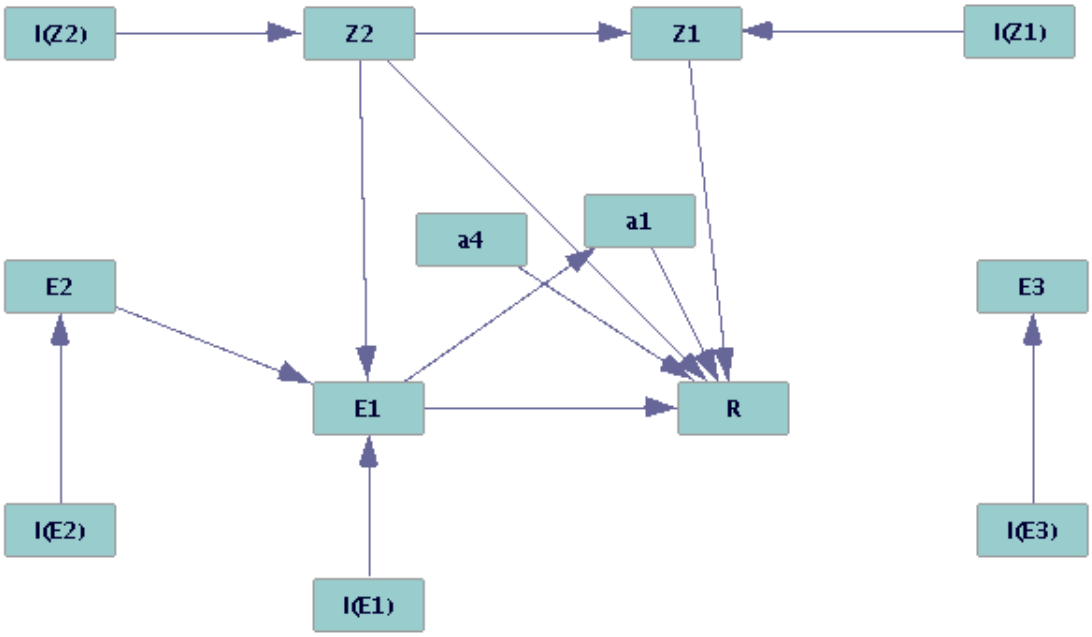
Suppose we find that for all $\gamma \setminus \{\gamma_1, \gamma_4\}$, $\gamma \neq 0$, for all η, $\eta = 0$, and at least for some combinations of ranges of values for each of the environmental variables, $\gamma_1, \gamma_4 \neq 0$. Then we can generate the following partial orders: $Z_2 \succ E_1$ and $Z_2 \succ Z_1 \succ R$. Step 3 is to determine the direct causal effects of the phenotypic instruments on the phenotypic variables. Following this, we quickly move to step 4. Because $Z_1 \succ R$ we can determine the direct causal effect of Z_1 on R for different ranges of values of variables in $\mathbf{E}$ and generate values of selection gradient $\alpha_1 = \dfrac{\gamma_1}{\beta_{I_{Z_1} Z_1}}$, for different values of γ_1 as previously

determined in step 2. We can now also assign Z to the $\alpha_1 \to R$ edge. Step 5 is to determine the direct causal effects of Z_2. The direct effect of Z_2 on Z_1 is $\alpha_2 = \frac{\gamma_2}{\beta_{I_{Z_2} Z_2}}$.

The direct effect of Z_2 on on E_1 is $\alpha_3 = \frac{\gamma_3}{\beta_{I_{Z_2} Z_2}}$. For concreteness, we now have the updated graph in figure 5.8. Now we move to step 6: Sort the partial orderings involving the environmental variables plus reproductive success. To do this we must perform the following set of standardized regressions:

$E_1 = \gamma_6 I_{E_2}$ and $R = \gamma_8 I_{E_2}$, for $Z_1, Z_2 = 0$ and $Z_2 = \eta_7 I_{E_2}$ and $Z_1 = \eta_8 I_{E_2}$ and $E_3 = \eta_9 I_{E_2}$
$R = \gamma_7 I_{E_1}$, for $Z_1, Z_2 = 0$ and $Z_2 = \eta_{10} I_{E_1}$ and $Z_1 = \eta_{11} I_{E_1}$ and $E_2 = \eta_{12} I_{E_1}$ and $E_3 = \eta_{13} I_{E_1}$
$Z_2 = \eta_{14} I_{E_3}$ and $Z_1 = \eta_{15} I_{E_3}$ and $E_1 = \eta_{16} I_{E_3}$ and $E_2 = \eta_{17} I_{E_3}$ and $R = \eta_{18} I_{E_3}$, for $Z_1, Z_2 = 0$

Suppose we find that for all $\gamma, \gamma \neq 0$ and for all $\eta, \eta = 0$. From this and the above, we can now construct the following partial orders: $Z_2 \succ \{Z_1, E_1\} \succ R$ and $E_2 \succ E_1 \succ R$. Step 7 is to determine the direct causal effects of the environmental variable instruments on their respective environmental variables. After this, step 8 involves determining the direct effects of the environmental variables without determining whether they are causes of selection. The direct effect of E_1 on R is $\alpha_7 = \frac{\gamma_7}{\beta_{I_{E_1} E_1}}$. The direct effect of E_2 on E_1 is $\alpha_6 = \frac{\gamma_6}{\beta_{I_{E_2} E_2}}$. Next is step 9. Because $Z_2 \succ ... \succ Z_1 \succ ... \succ R$, we can determine the direct

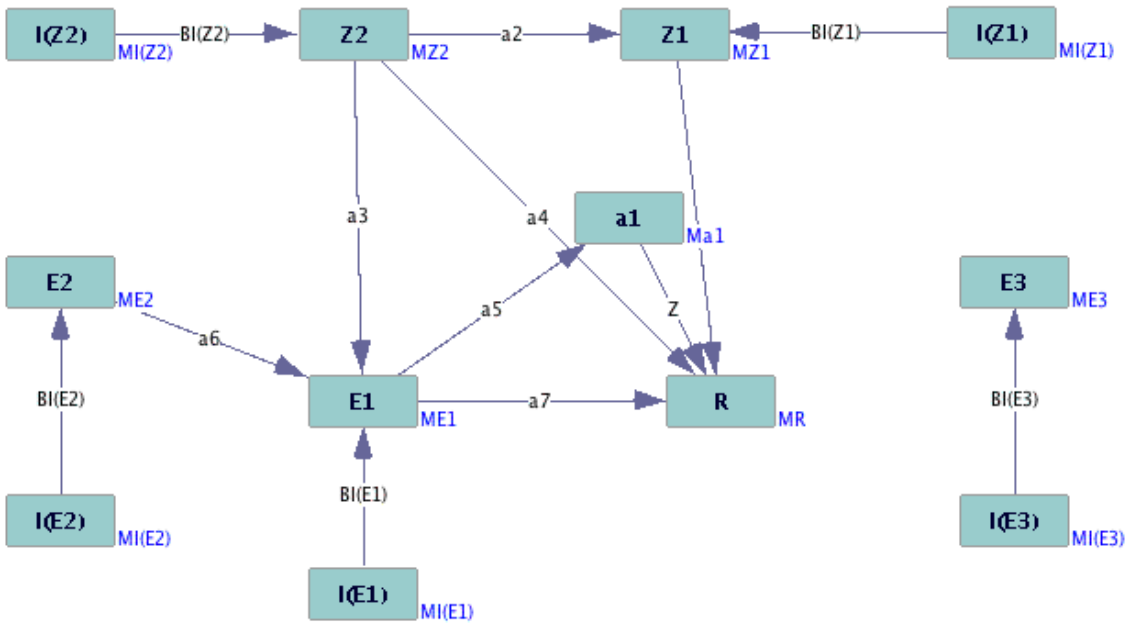


5.9. Updated Knowledge graph.

causal effect of Z_2 on R now. Perform the following standardized regression for different ranges of values of the environmental variables and for $Z_1 = 0$:

$$R = \gamma_4 I_{Z_2}, \text{ for } Z_1 = 0$$

Suppose we find that at least for some combination of values for the environmental variables the casual parameter is nonzero. Then we can generate $\alpha_4 = \frac{\gamma_4}{\beta_{I_{Z_2} Z_2}} - \alpha_3 \alpha_7$, for different values of γ_4. We can also now assign Z_2 to the edge $\alpha_4 \rightarrow R$. Step 10 is to



5.10. Returned Knowledge Graph with Causal Parameters.

determine which environmental variables cause selection on our phenotypic variables. To do this, we must perform the following set of standardized regressions:

$$\alpha_1 = \gamma_5 I_{E_1} \quad \text{and} \quad \alpha_4 = \eta_{19} I_{E_1}$$
$$\alpha_1 = \gamma_9 I_{E_2} \quad \text{and} \quad \alpha_4 = \eta_{20} I_{E_2}$$
$$\alpha_1 = \eta_{21} I_{E_3} \quad \text{and} \quad \alpha_4 = \eta_{22} I_{E_3}$$

Suppose we find that $\gamma_5, \gamma_9 \neq 0$ and find that for all $\eta, \eta = 0$. From this we can construct the following partial orders: $Z_2 \succ Z_1 \succ R$ and $Z_2 \succ E_1 \succ \alpha_1 \succ R$ and $E_2 \succ E_1 \succ \alpha_1 \succ R$. We also already know that both Z_2 and E_1 have direct effects on reproductive success.

Hence, we can determine the direct effect of E_1 on α_1, $\alpha_5 = \frac{\gamma_5}{\beta_{I_{E_1}E_1}}$. Variable E_1 is a direct cause of selection. We now have the updated knowledge graph in figure 5.9. The knowledge graph in figure 5.9 contains some variables that have only indirect causal effects on other variables. Step 11 of the pseudo-code involves testing whether these variables have direct causal effects on those variables. We have the following directed paths of length greater than one:

$Z_2 \to E_1 \to \alpha_1$ and $E_2 \to E_1 \to R$ and $E_2 \to E_1 \to \alpha_1$ and $E_2 \to E_1 \to \alpha_1 \to R$

We stipulated at the outset for ease that Z_2 has no direct causal influence on selection gradients. So, we can ignore the first directed path. Next we need to consider whether E_2 has a direct causal influence on α_1 or R. Suppose that from the last batch of standardized regressions $\frac{\gamma_9}{\beta_{I_{E_2}E_2}} = \alpha_6\alpha_5$. Then no direct edge from E_2 to the selection gradient on Z_1 is added. Variable E_2 is not a direct cause of selection but an indirect cause of selection. We can do the same with respect to its causal influence on reproductive success as well. Suppose that from step 6, $\frac{\gamma_8}{\beta_{I_{E_2}E_2}} = \alpha_6\alpha_7$. Then no direct edge from this environmental variable to reproductive success is added. The causal structure among the observed variables has been established. The output is figure 5.10.

To see why this works generally, consider the possibility of false positives (cf.

Eberhardt 2009 on interventions). Suppose that we infer a nonzero causal parameter even though it is the case that the causal parameter is zero. There are two possible reasons for this.

First, we might have inferred the nonzero causal parameter on the edge between two variables X_k and X_l because the variables satisfied the definition of a partial ordering over them. Yet, this is impossible because $X_k \rightarrow X_l$ if and only if $X_k \succ X_l$. If there is a directed path from I_{X_k} to X_l mediated by X_k, then by Pearl's d-separation, I_{X_k} and X_l are unconditionally associated. If they are unconditionally associated, then by Pearl's d-separation, they are causally connected (in some way). From the assumptions about instrumental variables, we know that it must be the case that $I_{X_k} \rightarrow X_k$ and we also know that there is no causal connection between I_{X_k} and X_l that does not "go through" X_k. Hence, if there is a causal connection, it must be such that there is a directed path from I_{X_k} to X_l mediated by X_k. So, we know by Pearl's d-separation and our assumptions about instrumental variables that $\mathrm{CORR}(I_{X_k}, X_l) \neq 0$ if and only if there is a directed path from the instrumental variable to X_l mediated by X_k. In addition, there cannot be any other measured variable X_m that mediates between X_k and X_l. For, if this were the case, then we would have found that I_{X_k} and X_m were unconditionally associated and that I_{X_m} and X_m were unconditionally associated. Yet, we explicitly took this into account in the definition of a partial ordering over variables with constraint

(PO2). Thus, it is impossible to get a nonzero causal parameter when it in fact is zero in this case.

There is a second reason why we might have gotten a false positive. We might have inferred the nonzero causal parameter on the edge between X_k and X_l because either we shut off the interactive effect X_k (with another variable X_m) has on X_l by regressing X_l on X_k, for $X_m = 0$ and found a nonzero regression estimate or because the unconditional association between X_k and X_l is not fully accounted for by their known indirect causal relationship(s). If only the former is the case and there is no indirect causal relationship besides the interactive one, then there is no new special problems (from what was discussed in the previous paragraph) as the interactive causal relationship has been shut off. If the latter or the former and the latter is the case, however, note that there must be at least one indirect path between X_k and X_l that is not "shut off." Assume, that something like the following causal graph is true:

$$X_k \rightarrow ... \rightarrow X_m \rightarrow ... \rightarrow X_l$$

It turns out that for all paths in a **POG**, all direct causal parameters can be determined. To see this, assume it is the case that all direct causal parameters can be determined for paths of length r. Now consider paths of length $r+1$ between X_k and X_l. We know all of the direct causal parameters along the path from applying the definition of a partial ordering over variables in the path. Thus, we know there is at least one directed path from X_k and X_l and we know the indirect causal effect of X_k on X_l. If we let the number of such

directed paths be p and let α_{ij} be the causal parameter j on path i, then this known causal effect is given by $\sum_i^p \prod_j^{r+1} \alpha_{ij}$ (Wright 1934). Yet, we do not know whether there is a direct causal effect between X_k and X_l. Happily, we can find this given what we know and our assumptions. We know the direct causal effect $\beta_{I_{X_k} X_k}$ of the instrumental variable on X_k. We know the total causal effect $\gamma_{I_{X_k} X_l}$ of the instrument of X_k on X_l. So, by linearity, the total causal effect of X_k on X_l is $\frac{\gamma_{I_{X_k} X_l}}{\beta_{I_{X_k} X_k}}$. So, given that we know the indirect causal effect and the total causal effect: $\frac{\gamma_{I_{X_k} X_l}}{\beta_{I_{X_k} X_k}} \neq \sum_i^p \prod_j^{r+1} \alpha_{ij}$ if and only if $X_k \rightarrow X_l$ in addition to their indirect causal relationship or $\frac{\gamma_{I_{X_k} X_l}}{\beta_{I_{X_k} X_k}} = \sum_i^p \prod_j^{r+1} \alpha_{ij}$ if and only if the two variables are merely indirectly linked. Then the causal parameter on that edge is determined by $\frac{\gamma_{I_{X_k} X_l}}{\beta_{I_{X_k} X_k}} - \sum_i^p \prod_j^{r+1} \alpha_{ij}$ because of our linearity assumption (Wright 1934). Hence, by mathematical induction over the lengths of paths between two variables, for all paths in a **POG**, all direct causal parameters can be determined (even when they are zero). Therefore, false positives are impossible here as well.

6. Summary and Possible Extensions

It is necessary for evolutionary biologists to develop a causally explicit methodology. In order to tell whether a trait has evolved by selection or drift, we must explicitly model the causes of reproductive success. Sometimes this also means explicitly modeling how selection on traits changes as a function of environmental variables. The main reason for needing to do this is because without doing so one cannot distinguish between, on the one hand, when selection is fluctuating and, on the other hand, when drift merely makes it look as though selection is fluctuating. When evolutionary biologists have been interested in fluctuating selection, they have sometimes failed to take steps to avoid such conflation. Rather than explicitly model the causal influence of environmental variables, selection differentials and selection gradients are estimated separately for locations in spacetime. In this chapter, I have argued that extending the use of selection differentials or mere regression-based selection gradients to contexts where we have data on specific environmental variables that might cause selection to fluctuate is not advisable. Instead, evolutionary biologists need an explicitly causal methodology to discover the causes of reproductive success and natural selection. I have offered the first steps toward that end within an instrumental variable framework.

There are many ways in which the causal methods here can be extended. One extension would be to construct a method for discovering instrumental variables for future research. To do this, we need one more definition:

A tier ordering (Spirtes et al. 2000) is an ordering of mutually exclusive and exhaustive sets of measured variables in a causal graph such that tier one contains all of the exogenous variables and, for all tiers $t > 1$, X_t is in tier t if and only if

(T1) there is a X_{t-1} in tier $t-1$ such that $X_{t-1} \rightarrow X_t$ on the graph and

(T2) there is no X_u in tier $u \geq t$ such that X_u is a direct or indirect cause of X_t.

Given this definition, after we have returned some graph like that found in the last section, we can discover unmeasured common causes (cf. Eberhardt 2009), and with the discovery of unmeasured common causes, one can, in turn, identify whether some variable in $\mathbf{S}$ in the present study might be used as instrumental variables in an analogous future study. For example:

1. Given the output from the pseudo-code in the last section and a set of hypotheses of which pairs of variables have unmeasured common causes, let $\mathbf{T} = T_1 \succ T_2 \succ ... \succ T_t \succ ... \succ T_w$ be a tier ordering over the measured variables (including, selection gradients) with w tiers. For all pairs of variables $X_k, X_l \in T_t$ (beginning with T_1), test the unconditional association between the considered pair.
2. Compare the known association with the just-tested unconditional association in step 1. If they are equal, then there is no unmeasured common cause between the considered pair. If they are unequal, then there is at least one unmeasured common cause for the unaccounted unconditional association.
3. Next, for tier $u > t$ (beginning with T_{t+1}),

A. For all pairs of variables X_t, X_u such that $X_t \in T_t$ and $X_u \in T_u$, test the unconditional association between the considered pair.

B. Compare the known association with the just-tested unconditional association in step 3A. If they are equal, then there is no unmeasured common cause between the considered pair. If they are unequal, then there is at least one unmeasured common cause for the unaccounted unconditional association.

C. Return to step 3A with tiers t and $u+1$ until and including tier $u = w$.

4. Return to step 1 with tier $t+1$ until and including tier $t = w$.

After running this, we might discover there is no unmeasured common cause between any of the variables in the previous section and variable E_2. Given what we have discovered at this point, then, we have the basis for arguing for the use of a variable like E_2 in an analogous variable set for a variable like E_1 in future research.

Another possible extension takes us to group selection contexts (see e.g., Chapter 4). In the group selection literature, contextual analysis is essentially the same as the Lande and Arnold (1983) regression-based approach to selection gradients. It is arguable that an instrumental variable approach to this area of research is superior. For example, regression-based approaches cannot make important distinctions between indirect causal influence and causal connections merely occurring from common causes whereas an instrumental variable approach can. Still, extending the instrumental variable approach on offer here to group selection contexts is not so straightforward. This is because of the

problem of tautologous connections. In particular, simply supplementing regression with instruments in contexts of group selection is unwise because individual level variables and variables measuring something about the types one's group or neighboring peers are will often be mathematically related and, so, not necessarily genuinely causally related. This is because, in those methods, every unit who is a focal individual is also a group or neighborhood member in the data set, so mathematical relations abound. In order to infer non-tautologous causal relationships when demographic conditions cause reproductive success (see e.g., Chapter 4), we will need to supplement the instrumental variable approach.

CONCLUSION

1. Summing Up

Most of evolutionary genetics since the Modern Synthesis has emphasized the concept of fitness and statistical formulae when cashing out natural selection, either conceptually or mathematically. Consequently, there has been an unfortunate deemphasis on clear causal reasoning and, in particular, the fact that at bottom natural selection requires that some trait variable causes reproductive success. As with any foregrounding and backgrounding in science this has come with advantageous and costs. I have focused on the costs in this dissertation, and I hope that doing so—with aid of the interventionist framework of causation—has shed light on old conceptual problems and provided new solutions to those traditional problems. I have attempted to make headway on a number of conceptual problems: The dissertation provided a new theory of adaptations and illuminated the implications of such an account (Chapter 2); it provided new conceptions of selection and drift by examining a case of a novelty evolved by (a particular form of) niche construction (Chapter 3); and it argued for the importance of sorting selection phenomena causally when diagnosing the levels of selection (on one view of multilevel selection) and argued against model pluralism (Chapter 4). I also took some initial steps toward constructing an explicit causal methodology for inferring the causes of natural selection (Chapter 5). The dissertation has advanced a theory of natural selection that is fundamentally causal in nature. In addition, it has opened up further avenues of research, which one could pursue in the future.

2. Future Directions

In the last section of the previous chapter, I discussed some possible extensions of my methodological work. Here I focus on how one might extend my more theoretical and conceptual work. First, one could explore whether the direction of causality between traits and environments matters to evolutionary outcomes. It might be thought that whether traits cause environments or the converse or both matters little as far as evolutionary outcomes are concerned. If there is some probabilistic association between traits and environments, then this can be incorporated into statistical models and evolutionary trajectories can be predicted. I doubt that this is true generally. In order to study this issue one might intend to study phenotypic plasticity (Pigliucci 2001) and include it into some of the generative models discussed in the dissertation. I predict that this kind of study will also lead to conceptual clarification regarding how to interpret cases with causal connections among the causes of reproductive success. In particular, it should shed light on particular cases of evolution by natural selection such as those related to cancer where both niche construction and phenotypic plasticity seem to be at play at multiple levels of biological organization (Lean and Plutynski 2016).

Second, one might want to explore coevolution on two fronts: the coevolution of spatially related species on small and large geographical scales (Thompson 2005) and the coevolution of populations and their demographic structure in order to make headway on problems having to do with evolutionary transitions (Maynard Smith and Szathmáry 1995; Michod 2000). Much of the former work is inherently spatial, but the current tools do not adequately reflect this and much of the latter work seems ripe for conceptual

clarification and clear causal reasoning. The coevolutionary process can be modeled in the following way. Say that a single population of types or two populations of species initially exists on the $\mathbb{Z}^d$ integer lattice and assume that there is an infinite connected subgraph represented by the vertex set $\mathbf{V}$. Whether an individual of some type or species invades this subgraph depends on whether one of the subgraph's vertices is in her set of $2d$ nearest neighboring spaces. The dynamics of the sites outside of the subgraph, then, are locally determined. However, within the subgraph—because all sites within it are connected—the dynamics are globally determined (e.g., by the replicator equation). Finally, suppose that depending on whether there are more or less of some type within $\mathbf{V}$, the subgraph shrinks (if the type is harmed by being in the subgraph) or expands (if the type benefits from being in the subgraph). Given this set-up, one can either examine how two species coevolve over a landscape where dispersal ranges change or examine the evolutionary transitions among types and demographic structures together. Moreover, one can also examine these coevolutionary processes on separate time scales, with types or species evolving according to one scale and landscape or demographic structure evolving according to another.

Third, the coevolutionary processes described above can also be examined with the addition of niche construction or phenotypic plasticity. Some analogous previous work by Brown (1990)—extending the work of Rosenzweig (1987)—found that in a two type environment, there are three scenarios possible given a population consisting of two specialists and one generalist phenotypes: (1) the two specialists coexist; (2) the generalist evolves to fixation; and (3) one of the specialists and the generalist coexist. The

first case occurs when there is a cost to being a generalist, or alternatively, when there is no cost to being a specialist. The second case occurs when there is no cost to being a generalist, or alternatively, when there is a cost to being a specialist. The final case occurs when there is an asymmetric interactive causal influence of the environmental variable on reproductive success or there is an asymmetry in the frequency of different environments. In my envisioned future work, the two environments would correspond to the two types of demographic structure. Given this, it is interesting to consider whether scenarios (1-3) occur given different causal influences of type and demographic structure, where the landscape or demographic structure does not also evolve, and consider what happens with a dynamic evolving landscape or demographic structure.

Fourth and finally, further research might turn to issues on cultural evolution, its foundations and analysis. In particular, I am interested in studying variants of Lewis's (1969) signaling games. Suppose a population of early hominids lives in an environment with some number n of objects, and suppose there is some repertoire of some number m of signals. Then, agents can be characterized by a $n \times m$ probability matrix producing a word given the sight of an object and a $m \times n$ matrix producing the idea of an object given a word. When two random individuals attempt communication, the speaker sees an object and produces a signal for the hearer, the hearer hears the signal and produces an idea related to an object, both get at least one chance in each role, and a payoff (i.e., fitness) is rewarded to the individuals in proportion to the success of their communication. The language game continues again in future generations after parents die with offspring learning the language of their parent before parental death. What

happens at the end of this game is that a single (perhaps sub-optimal) language evolves in the population. Though—to my knowledge—there has not been any rigorous analysis on the trajectory to this state given different sorts of "learning periods" of offspring, populations sizes, and demographic structures. Indeed, on demographic structure, what I have just described corresponds to a completely connected population. Yet, many social structures in humans are not, of course, like this. Extending the work in this dissertation to this aspect of cultural evolution is likely to be both enormously interesting and philosophically rich.

REFERENCES

Anderson, Wes (in press). Some adaptations were not positive causal factors for reproductive success. To appear in *Philosophy of Science*.

Angrist, Joshua D. and Jorn-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Antonakis, John, Samuel Bendahan, Philippe Jacquart, and Rafael Lalive (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly* **21**: 1086-1120.

Aspi, Jouni, Anne Jäkäläniemi, Juha Tuomi, and Pirkko Siikamäki (2003). Multi-level phenotypic selection on morphological characters in a metapopulation of *Silene Tatarica*. *Evolution* **57**: 509-517.

Beatty, John. (1984). Chance and natural selection. *Philosophy of Science* **51**: 183-211.

Brandon, Robert. (1990). *Adaptation and Environment*. Princeton, NJ: Princeton University Press.

Brandon, Robert and H. Frederik Nijhout (2006). The empirical nonequivalence of genic and genotypic models of selection: A (decisive) refutation of genic selectionism and pluralistic genic selectionism. *Philosophy of Science* **73**: 277-297.

Brown, Joel S. Habitat selection as an evolutionary game. *Evolution* **44**: 732-746.

Burnham, Kenneth P. and David R. Anderson (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretical Approach*. Dordrecht: Springer.

Cameron, A. Colin and Pravin K. Trivedi (2009). *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Damuth, John and I. Lorraine Heisler (1988). Alternative formulations of multilevel selection. *Biology and Philosophy* **3**: 407-430.

Didelez, Vanessa, Sha Meng and Nuala A. Sheehan (2010). Assumptions of IV methods for observation epidemiology. *Statistical Science* **25**: 22-40.

Eberhardt, Frederick (2009). *Causation and Intervention*. PhD dissertation, Carnegie Mellon University.

Endler, John (1983). Natural selection on color patterns in *Poecilia reticulata*. *Evolution* **34**: 76-91.

Endler, John. (1986). *Natural Selection in the Wild*. Princeton, NJ: Princeton University Press.

Evilsizor, Stephen and Nicolas Lanchier (2015). Evolutionary games on the lattice: Death-brith updating process. arXiv preprint arXiv: 1506.07233.

Gibbs, H. Lisle, and Peter R. Grant (1987). Oscillating selection on Darwin's finches. *Nature* **327**: 511-513.

Gildenhuys, Peter. (2009). An explication of the causal dimension of drift. *British Journal for the Philosophy of Science* **60**: 521-555.

Gillespie, John H. (1974). Natural selection for within-generation variance in offspring number. *Genetics* **76**: 601-606.

Glymour, Bruce. (2011). Modeling environments: Interactive causation and adaptations to environmental conditions. *Philosophy of Science* **78**: 448-471.

Glymour, Bruce and Christopher French (unpublished). Causation, group selection, and equivalence.

Godfrey-Smith, Peter (2008). Varieties of population structure and the levels of selection. *The British Journal for the Philosophy of Science* **59**: 25-50.

Goodnight, Charles J., James M. Schwartz, and Lori Stevens (1992). Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. *American Naturalist* **140**: 743-761.

Heckman, James L. (2000). Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics* **115**: 45-97.

Heisler, I. Lorraine and John Damuth (1987). A method for analyzing selection in hierarchically structured populations. *American Naturalist* **130**: 582-602.

Hereford, Joe, Thomas F. Hansen, and David Houle (2004). Comparing strengths of directional selection: How strong is strong? *Evolution* **58**: 2133-2143.

Heywood, James (2010). Explaining patterns in modern ruminant diversity: Contingency or constraint? *Biological Journal of the Linnean Society* **99**: 657-672.

Hutchinson, G. Evelyn (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* **22**: 415-427.

Kerr, Benjamin and Peter Godfrey-Smith (2002). Individualist and multi-level perspectives on selection in structured populations. *Biology and Philosophy* **17**: 477-517.

Kettlewell, Bernard (1973). *The Evolution of Melanism*. Oxford: Claredon Press.

Kingsolver, Joel G., Hopi E. Hoekstra, Jon M. Hoekstra, David Berrigan, Sacha N. Vigniere, C.E. Hill, Anhthu Hoang, P. Gilbert, and Peter Beerli (2001). The strength of phenotypic selection in natural populations. *The American Naturalist* **157**: 245-261.

Laland, Kevin N., Tobias Uller, Marcus W. Feldman, Kim Sterelny, Gerd B. Müller, Armin Moczek, Eva Jablonka, and John Odling-Smee (2015). The extended evolutionary synthesis: Its structure, assumptions, and predictions. *Proceedings of the Royal Society B* **282**: DOI: 10. 1098/rspb.2015.1019

Lanchier, Nicolas (2015). Evolutionary games on the lattice: Payoffs affecting birth and death rates. *Annual Applied Probability* **25**: 1108-1154.

Lande, Russell and Steven J. Arnold (1983). The measurement of selection on correlated characters. *Evolution* **37**: 1210-1226.

Laubichler, Manfred and Jürgen Renn (2015). Extended evolution: A conceptual framework for integrating regulatory networks and niche construction. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **324**: 565-577.

Lean, Christopher and Anya Plutynski (2016). The evolution of failure: Explaining cancer as an evolutionary process. *Biology and Philosophy* **31**: 39-57.

Lennox, James and Bradley Wilson (1994). Natural selection and the struggle for existence. *Studies in History and Philosophy of Science* **25**: 65-80.

Lewis, David (1969). *Convention: A Philosophical Study*. Harvard University Press.

Lewontin, Richard. (1970). The units of selection. *Annual Review of Ecology and Systematics* **1:** 1-18.

Lewontin, Richard. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics* **26**: 400-411.

Lewontin, Richard. (1985). Adaptation. In R. Levins and R. Lewontin, *The Dialectical Biologist*. Cambridge, MA: Harvard University Press, 65-84.

MacColl, Andrew D.C. (2011). The ecological causes of evolution. *Trends in Ecology & Evolution* **26**: 514-522.

Maynard Smith, John and Eors Szathmáry (1995). *The Major Transitions in Evolution*. Oxford University Press.

Mayr, Ernst (1961). Cause and effect in biology. *Science* **134**: 1501-1506.

Michod, Richard E. (2000). *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton, NJ: Princeton University Press.

Mills, Susan. and John Beatty (1979). The propensity interpretation of fitness. *Philosophy of Science* **46**: 263-286.

Millstein, Roberta. (2002). Are random drift and natural selection conceptually distinct? *Biology and Philosophy* **17**: 33-53.

Mitchell-Olds, Thomas and Ruth G. Shaw (1987). Regression analysis of natural selection: Statistical inference and biological interpretation. *Evolution* **41**: 1149-1161.

Müller, Gerd B. and Stuart A. Newman (2005). The innovation triad: An EvoDevo agenda. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **304**: 487-503.

Neuhauser, Claudia and Stephen W. Pacala (1999). An explicitly spatial version of the Lotka-Volterra model with interspecific competition. *The Annals of Applied Probability* **9**: 1226-1259.

Nowak, Martin A. (2006). *Evolutionary Dynamics*. Cambridge, MA: Harvard University Press.

Ohtsuki, Hirashi, Christoph Hauert, Erez Lieberman, and Martin A. Nowak (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**: 502-505.

Okasha, Samir. (2004). Multi-level selection, covariance, and contextual analysis. *British Journal for the Philosophy of Science* **55**: 481-504.

Okasha, Samir. (2006). *Evolution and the Levels of Selection*. Oxford: Oxford University Press.

Page, Karen M. and Martin A. Nowak (2002). Unifying evolutionary dynamics. *Journal of Theoretical Biology* **219**: 93-98.

Pearl, Judea. (2000). *Causality*. Cambridge: Cambridge University Press.

Pence, Charles and Grant Ramsey (2013). A new foundation for the propensity interpretation of fitness. *The British Journal for the Philosophy of Science* **64**: 851-881.

Pfeifer, Jessica (2005). Why selection and drift might be distinct. *Philosophy of Science* **72**: 1135-1145.

Pigliucci, Massimo (2001). *Phenotypic Plasticity*. Oxford University Press.

Pigliucci, Massimo. and Jonathan Kaplan (2006). *Making Sense of Evolution: The Conceptual Foundations of Evolutionary Biology.* Chicago, IL: University of Chicago Press.

Price, George. (1970). Selection and covariance. *Nature* **227**: 520-521.

Price, George. (1972). Extension of covariance selection mathematics. *Annals of Human Genetics* **35**: 485-490.

Price, Trevor D., Peter R. Grant, H. Lisle Gibbs, and P.T. Boag (1984). Recurrent patterns of natural selection in a population of Darwin's finches. *Nature* **309**: 787-789.

Quinn, Gerry P. and Michael J. Keough (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge: Cambridge University Press.

Rice, Sean H. (2004). *Evolutionary Theory*. Sunderland, MA: Sinauer.

Robertson, Alan (1966). A mathematical model of the culling process in dairy cattle. *Animal Production* **8**: 95-108.

Rosenzweig, Michael L. (1987). Habitat selection as a source of biological diversity. *Evolutionary Ecology* **1**: 315-330.

Schluter, Dolph (1988). Estimating the form of natural selection on a quantitative trait. *Evolution* **42**: 849-861.

Shanahan, Timothy (1989). Beatty on chance and natural selection. *Philosophy of Science* **56**: 484-489.

Sheehan, Nuala A., Vanessa Didelez, Paul R. Burton, and Martin D. Tobin (2008). Mendelian randomization and causal inference in observation epidemiology. *PLoS Med* **5**: e177.

Shipley, Bill (2000). *Cause and Correlation in Biology*. Cambridge: Cambridge University Press.

Siepielski, Adam M., Joseph D. DiBattista, and Stephanie M. Carlson (2009). It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecology Letters* **12**: 1261-1276.

Siepielski, Adam M., Kiyoko M. Gotanda, Michael B. Morrissey, Sarah E. Diamond, Joseph D. DiBattista, and Stephanie M. Carlson (2013). The spatial patterns of directional phenotypic selection. *Ecology Letters* **16**: 1382-1392.

Sober, Elliott. (1984). *The Nature of Selection*. Chicago, IL: University of Chicago Press.

Sober, Elliott and David S. Wilson (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Sober, Elliott and David, S. Wilson (2002). Perspectives and parameterizations: Commentary on Benjamin Kerr and Peter Godfrey-Smith's "Individualist and multi-level perspectives on selection in structured populations." *Biology and Philosophy* **17**: 529-537.

Sokal, Robert R. and F. James Rohlf (2012). *Biometry*. New York, NY: W.H. Freeman & Company.

Spirtes, Peter., Clark. Glymour, and Richard. Schienes (2000). *Causation, Prediction, and Search*. Cambridge, MA: MIT Press.

Sterelny, Kim (2009). Novelty, plasticity, and niche construction: The influence of phenotypic variation on evolution. In Anouk Barberousse, Michel Morange, and Thomas Pradeu (eds.), *Mapping the Future of Biology.* Springer Netherlands, 93-110.

Sterelny, Kim and Philip Kitcher (1988). The return of the gene. *The Journal of Philosophy* **85**: 339-361.

Thompson, John N. (2005). *The Geographic Mosaic of Coevolution*. Chicago: University of Chicago Press.

Wade, Michael J. (1985). Soft selection, hard selection, kin selection, and group selection. *American Naturalist* **125**: 61-73.

Wade, Michael J. and Susan Kalisz (1990). The causes of natural selection. *Evolution* **44**: 1947-1955.

Wagner, Günter P. (2014). *Homology, Genes, and Evolutionary Innovation.* Princeton University Press.

Walker, Jeffrey A. (2014). The effect of unmeasured confounders on the ability to estimate a true performance or selection gradient (and other partial regression coefficients). *Evolution* **68**: 2128-2136.

Weinberger, Naftali (2011). Is there an empirical disagreement between genic and genotypic selection models? A response to Brandon and Nijhout. *Philosophy of Science* **78**: 225-237.

Wilson, David S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences, USA* **72**: 143-146.

Wilson, Robert A. (2003). Pluralism, entwinement and the levels of selection. *Philosophy of Science* **70**: 531-552.

Woodward, James. (2003). *Making Things Happen*. Oxford: Oxford University Press.

Wright, Sewall (1934). The method of path coefficients. *Annals of Mathematical Statistics* **5**: 161-215.