Brit. J. Phil. Sci. **57** (2006), 627–653

Natural Selection as a Population-Level Causal Process

Roberta L. Millstein

ABSTRACT

Recent discussions in the philosophy of biology have brought into question some fundamental assumptions regarding evolutionary processes, natural selection in particular. Some authors argue that natural selection is nothing but a population-level, statistical consequence of lower-level events (Matthen and Ariew [2002]; Walsh et al. [2002]). On this view, natural selection itself does not involve forces. Other authors reject this purely statistical, population-level account for an individual-level, causal account of natural selection (Bouchard and Rosenberg [2004]). I argue that each of these positions is right in one way, but wrong in another; natural selection indeed takes place at the level of populations, but it is a causal process nonetheless.

1 Introduction

Recent discussions of natural selection have given conflicting answers to a pair of questions: first, is natural selection a causal process or is it a purely statistical aggregation? And second, is natural selection at the population level or at the level of individuals? Denis Walsh, Tim Lewens and André Ariew ([2002]) and Mohan Matthen and Ariew ([2002]) (hereafter, when I am referring to the two papers together, WALM) have argued that natural selection is *purely statistical* and on the *population level*, whereas Frédéric Bouchard and Alex Rosenberg ([2004]) (hereafter, BR) argue that natural selection is *causal* and on the *individual level.* In this essay, I argue for a

doi:10.1093/bjps/axl025
Advance Access published on November 14, 2006

third logical possibility: natural selection is indeed a *causal* process, but it operates at the *population* level.

The issues at stake in this debate will be fleshed out in more detail in the discussion below. However, at the outset it is important to distinguish this debate from the more well-known debates over the units (or levels) of selection; unfortunately, similar terminology makes this distinction difficult. One obvious difference between the debates is that the units of selection debates assume a causal basis for natural selection, whereas the existence of selection's causal basis is one of the points of contention in the debate at hand. With regard to the issue of levels, however, the difference between the two debates can best be seen by the following example. Suppose that a person believes, in terms of the levels of selection debate, that selection (either sometimes or always) acts on organisms. Such a person still might ask whether that selection process was acting on individual organisms or populations of organisms.[1] It is the latter question that concerns us here, and indeed, throughout this paper I have, for the sake of simplicity, assumed organismic selection, although I believe that similar arguments can be made for other units of selection. Now, it may turn out that the present debate has interesting consequences for the debates over the units of selection, but I will not explore such consequences here.

Further terminological confusion arises because of the distinction being made between an individual and a population. Indeed, one might wish to speak of an 'individual' population rather than a class or type of population. However, the reader should understand that, in what follows, 'individual' refers to an individual organism, whereas a population refers to a particular spatiotemporal collection of organisms.

Some philosophers might object to the very idea of population-level causality. In response to this concern, I will argue in Section 2.1 that anyone who accepts the reality of frequency-dependent selection is already committed to population-level causality. Moreover, in Section 2.2 I will argue that population-level causality is consistent with three commonly accepted accounts of causality; thus, population-level causality is at least of no more philosophical concern than they are. Turning to positive support for my claim that natural selection is a causal process that operates at the level of populations, I will need to make the case for both natural selection as 'causal' and natural selection as 'population-level.' I will take up the former issue in Section 3 and the latter issue in Section 4. Recall, as I noted above, that

[1] Frédéric Bouchard ([personal communication]) has suggested that in the units of selection debate, when someone believes that selection acts on organisms, they believe that it acts on individual organisms; however, part of my point is that the question at hand, whether individual organisms or populations of organisms are acted upon, has been glossed over in the levels of selection debate.

I agree with BR, but disagree with WALM, on the question of whether natural selection is causal; thus, Section 3 responds to WALM's claims.[2] On the other hand, I agree with WALM, but disagree with BR, on the question of whether natural selection is population level; thus, Section 4 responds to BR's claims. Both Sections 3 and 4 involve an examination of Nathan Rank's and Elizabeth Dahlhoff's studies of the montane willow leaf beetle (in particular, Rank [1992]; Dahlhoff and Rank [2000]; Rank and Dahlhoff [2002]), on the assumption that it is important not only to examine biological theory, but also to consider how that theory is applied in an actual case.

2 A brief justification of population-level causality

Some would argue against the very idea of population-level causality. While I cannot hope to fully persuade such individuals in a paper of this length—indeed, my response here is at best partial—my aim in this section is to provide a fuller account of the philosophical arguments that justify population-level causality for natural selection specifically.

2.1 Frequency-dependent selection

Anyone who accepts the full range of processes modeled by population genetics is already forced to acknowledge at least one instance in which natural selection is a population-level causal process: frequency-dependent selection.[3] Frequency-dependent selection occurs when 'the fitness of a genotype (or of an allele) is affected by its frequency within the population' (Futuyma [1986], p. 166). In some cases, a genotype is fitter when it is rare (negative frequency-dependence); in other cases, a genotype can be fitter when it is common (positive frequency-dependence). As an example of negative frequency-dependence, consider a population in which a prey species mimics a poisonous organism of a different species (e.g. among butterflies). In such a situation, being a mimic confers higher fitness when it is rare than when it is common, because the predator (e.g. a bird) is more easily 'fooled' into avoiding the mimic when mimics are rare. As an example of positive frequency-dependence, consider a population where there are different genotypes, all poisonous, with different coloration patterns. In this case, an organism is fitter if it is of a common genotype because it is more likely that the predator will have encountered its type, and thus it will avoid the

[2] Others have developed critiques of WALM's position (see, e.g. Stephens [2004]; Rosenberg and Bouchard [2005], Shapiro and Sober [forthcoming]). Thus, I will focus primarily on arguing for my alternative position.

[3] Two instances, actually: a similar argument can be made for density-dependent selection, where the fitness of a genotype varies with the size of the population.

organism in question. In the discussion that follows, I will focus on negative frequency-dependence, although the same arguments can be made for positive frequency-dependence.

Frequency-dependent selection is a clear-cut case of population-level selection because the outcome of selection (the change in gene or genotype frequencies from one generation to the next) is determined by a population-level parameter: the frequency of genotypes within a population. (This will be discussed further below). Consider the case of negative frequency-dependence discussed in the previous paragraph. In a population with few mimics, mimics will be favored and thus tend to increase in numbers in subsequent generations, whereas in a population with many mimics, the number of mimics will increase more slowly, or perhaps even decrease, in subsequent generations.

Frequency-dependent selection is believed to be quite common; Futuyma ([1986], p. 166) remarks that 'it is likely that there is a frequency-dependent component in virtually all selection that operates in natural populations, for interactions among members of a population affect the selective advantage of almost all traits, and such interactions usually give rise to frequency-dependent effects.' So, at least in this sense, it may be the case that most instances of selection exhibit population-level causation, and the onus is on anyone who would try to deny the possibility of population-level causation altogether to produce an explanation for the phenomenon of frequency-dependent selection. Nonetheless, one still might have questions as to whether selection more generally exhibits population-level causation; perhaps we can use the clear-cut case of population-level causation in frequency-dependent selection to illuminate causation in selection more generally.

My argument here is relatively simple. First, some terminology. A population can be defined as 'a group of conspecific organisms that occupy a more or less well-defined geographic region and exhibit reproductive continuity from generation to generation' (Futuyma [1986], pp. 554–5). Even though a population is composed of individual organisms, it is important to distinguish between properties that apply to *individual organisms* and properties that characterize *the relationships among organisms*—that is, properties that apply to populations. For example, individual organisms have properties such as color, shape and length. Populations, on the other hand, have properties such as size (defined as the number of individuals), frequency (defined as the proportion of individuals of one type or another) and growth rate (defined as the rate of change in the number of individuals in the population). Thus, in a sense, population-level properties are properties that arise only given the collection and interaction of individuals. Nonetheless, a property can apply to a population without applying to the individuals in the population. For example, the growth rate of a population could be increasing

while the growth rate of an individual organism within the population remains constant (and vice versa).

In pointing out that frequency-dependent selection involving mimicry exhibits *causality*, I am implying that '*C* is a (probabilistic) cause of *E*', where *C* = the frequency of a mimic in a population and *E* = a change in frequency of the mimic in subsequent generations. More specifically, when *C* = a mimic's rarity in the population, then *E* = an increase in the frequency of mimics in subsequent generations, whereas when *C* = a mimic's commonality in the population, then *E* = a decrease in the frequency of mimics in subsequent generations. Furthermore, the reason that frequency-dependent selection involving mimicry exhibits *population-level* causality is that the 'cause', the frequency of mimics in the initial population, is a property of the population. I can see no way to construe mimic frequency, which is the proportion of mimics to non-mimics in the population, as the property of one individual; it is the property of all of the individuals taken together (i.e. of the population as a whole). Furthermore, the 'effect', a change in frequency over subsequent generations, is similarly a property of the population (rather than the individual)—we are not discussing increases and decreases in particular individuals, which could mean only their gain or loss of weight. The invocation of properties of a population as a cause and properties of a population as an effect implies that we have, in frequency-dependent selection, a case of population-level causation.

In this essay, I will argue that just as frequency-dependent selection invokes population-level causes and population-level effects, so does selection in general. In particular, in Section 4 I will argue that, for the selection process in general, the 'cause' is variation in the population—more specifically, heritable differences in physical survival and/or reproductive abilities—a property of the population rather than of the individual.[4] Similarly, I will

[4] One might use a similar line of argument to show that selection in general exhibits population-level causation by claiming that the *size* of the population, a property of the population, is a cause involved in the selection process. That is, one might claim that selection predominates in large populations, whereas drift predominates in small populations. And it is true that with these phenomena, the size of the population can be said to play a causal role. For example, consider two populations undergoing random drift in identical environments, where the populations are identical in their genotype frequencies and differ only in their sizes. In this case, we would expect greater fluctuations in gene frequencies from one generation to the next in the smaller population than in the larger (just as small samples tend to be more unrepresentative than large samples). However, while I think there may be some merit to such an argument (implicit, perhaps, in an article by Reisman and Forber [forthcoming]), by focusing on population size it does not pick out a characteristic that is *definitive* of selection or drift. (Patrick Forber [personal communication] tells me that the goal of their essay is to determine the causal influence of founding population size, and not to define drift). Selection and drift need to be defined as processes of discriminate and indiscriminate sampling, respectively (Millstein [2002], [2005]), and discriminate sampling (selection) can occur in small populations, just as indiscriminate sampling (drift) can occur in large ones. Moreover, as Stephens emphasizes ([2004], [personal communication]),

argue that the 'effect' is differences in reproductive success, a property of the population rather than the individual. Thus, if one accepts frequency-dependent selection as population-level causation, one ought to accept selection in general as population-level causation as well.

2.2 Accounts of causation

But, causation in what sense? One problem that we run into immediately is the well-known lack of agreement on the nature of causation itself. Yet, I do not think we need to settle this issue here. Selection exhibits population-level causation in accordance with a number of different accounts of causation. Consider these three—a counterfactual account (e.g., Lewis [1973]), a manipulability account (e.g., Woodward [2003]) and a controlled experiment account (e.g., Dupré [1984]; Giere [1984]).[5] For each of the three cases, $C =$ heritable differences in physical characteristics (yielding differences in survival and/or reproductive abilities) and $E =$ differences in reproductive success:

1. Counterfactual account. On a counterfactual account, 'C is a cause of E' can be explained in terms of counterfactual conditionals of the form 'If C had not occurred, E would not have occurred.' Selection supports counterfactuals; if there were no heritable differences in physical characteristics among the organisms in the population, then there would be no differences in reproductive success (i.e. there would be nothing to be selectively favored or disfavored, as all the organisms would be of the same genotype).

2. Manipulability account. On a manipulability account (in a broad sense), if you can systematically manipulate C to bring about a change in E, then C is a cause of E. Indeed, if we were to change the heritable differences in physical

we must distinguish the expected outcome of drift from the actual outcome of drift; the former is solely a function of comparative population size whereas the latter is not. Even though we always *expect* a smaller population to exhibit greater fluctuations than a larger one from one generation to the next, the outcome that *in fact occurs* may be different (the small population may exhibit smaller fluctuations than the large one from one generation to the next). (Stephens [2004] says that he is distinguishing between 'process' and 'product', but it would probably be more accurate to say that he is distinguishing between 'expectation' and 'product').

[5] This is not meant to be an exhaustive list. For example, Hitchcock ([1996]) gives an account of selection and drift in terms of probabilistic causality. Although he does not say so explicitly, in making a distinction between causation in terms of causal processes and interactions (a position that Hitchcock attributes to Salmon) and causation in terms of probabilistic causality, Hitchcock implies that the former applies to the individual level whereas the latter applies to the population level. Much of what I have said here is congenial to Hitchcock's account in particular and probabilistic theories of causality in general, although I have, in contrast with Hitchcock, spoken of natural selection as a causal process. As I will emphasize, however, this is a population-level causal process (in the case of selection, a process of discriminate sampling), and so my use of the term differs from that which Hitchcock attributes to Salmon. (Thanks to Patrick Forber for directing me to Hitchcock's essay).

characteristics of the organisms in the population, we would probably change their relative reproductive successes. For example (and here I use an example that I will develop in more detail below), if we had a population of beetles that varied in their abilities to withstand hot and cold temperatures (based on heritable physical characteristics), and we introduced a new beetle genotype into the population that was able to withstand a greater range of temperatures than any of the existing genotypes, we would expect that the relative reproductive successes of the other genotypes would decrease (as a consequence of the success of the new genotype and the limited resources available for all beetles).[6] Thus, selection satisfies a manipulability account of causation.

3. Controlled experiment account. On a controlled experiment account, causation is captured by the conditions of a controlled (in the ideal case, randomized) experiment; take two populations selected at random and introduce *C* into one of them. '*C* is a cause of *E*' if the frequency of *E* is different in the two populations. To use the beetle example again, imagine two populations of beetles in the same cold environment, where there is a random distribution of alleles at all loci except for one (or, more plausibly, the two populations do not differ in the distribution of alleles at other loci). Let us call this one differing locus the 'PGI locus.' In one population, there is polymorphism at the PGI locus, with the different PGI genotypes differing in their abilities to withstand cold temperatures. For example, suppose there are three genotypes, PGI 1-1, PGI 1-4 and PGI 4-4, with PGI 1-1 having the greatest ability to withstand cold temperatures. In the second population, suppose that there is also polymorphism at the PGI locus (I specify this in order to make the two populations as similar as possible), but that the population consists of different PGI genotypes that do not differ in their abilities to withstand cold temperatures. Let us call these genotypes PGI *i-i*, PGI *i-j* and PGI *j-j*.[7] In the first population, we would expect, over the course of generations, an increase of the PGI 1-1 genotype over the other genotypes, whereas in the second population, we would not expect a sustained increase of any particular PGI genotype (e.g. an increase of PGI *i-i* is just as likely as an increase of PGI *j-j*, and we will probably see a fluctuation in the two genotypes over time). Thus, selection satisfies a controlled experiment account of causation.

Of course, such a short discussion as I have given here cannot hope to be definitive; it is only meant to show the plausibility of population-level

[6] See Reisman and Forber ([forthcoming]) for an argument that drift satisfies the manipulability condition. Reisman and Forber argue that drift is likewise a population-level causal process; thus, the present paper is much in sympathy with their conclusions.

[7] Whereas the previously named PGI 1-1, PGI 1-4 and PGI 4-4 genotypes are real (and will be discussed in further detail below), PGI *i-i*, PGI *i-j* and PGI *j-j* are imaginary.

causality for selection. That is, given that the aforementioned population-level properties embodied by selection satisfy these three well-known accounts of causation, concerns about the nature of causation, at least, are not reasons to reject population-level causation in selection. However, a positive account of why we should *accept* population-level causation in selection is needed, so let us turn to the first part of my claim, namely that natural selection is a *causal process*.

3 The montane willow leaf beetle: a causal story

My discussion of Rank and Dahlhoff's studies of the montane willow leaf beetle in this section is meant to illustrate three points: (1) many biologists understand natural selection as a causal process;[8] (2) their understanding of natural selection as a causal process affects their methods and the conclusions that they draw; and (3) that statistical analysis alone is usually insufficient to demonstrate natural selection (a demonstration of the specific causal process in play is required).

These claims are at odds with those of WALM. For example, Matthen and Ariew ([2002], p. 79) state:

> [. . .] natural selection is a mathematical aggregate of individual events. This seems to imply that it is not a causally connected process.[9]

Similarly, Walsh, Lewens, and Ariew ([2002], p. 453) assert, 'Selection and drift are not forces acting on populations; they are statistical properties of an assemblage of ''trial'' events: births, deaths and reproduction.' I understand both of these papers to be claiming that natural selection is *purely* statistical; natural selection, in their view, is a mere statistical summary of lower-level causes, but it is not itself a cause or a causal process. However, their arguments commit two mistakes. First, as Stephens ([2004]) has argued, these authors conflate the concept of 'force' with the concept of 'cause.' At best, they have shown that natural selection is not a force, but this fails to demonstrate that it is not a causal process. This brings us to the second mistake;

[8] For other arguments supporting the claim that natural selection is a causal process, see Sober ([1984]), Hodge ([1987]), Millstein ([2002]), Stephens ([2004]) and Shapiro and Sober ([forthcoming]).

[9] Towards the end of Matthen and Ariew's essay, there seems to be a relaxation of their claim that there is a dichotomy between statistical trends and forces; Matthen and Ariew ([2002], p. 82) 'concede that stochastic causation occurs at the [population] level.' However, I do not know how to reconcile this claim with their earlier remarks, and so throughout this essay I will assume that their position does deny population-level causality. But if I am mistaken in this, our positions may in fact be closer than it appears initially. Nonetheless, in 'denying that process causation occurs at this level' (Matthen and Ariew [2002], p. 82), Matthen and Ariew seem to be embracing a concept of process causation that is 'strictly law governed', and admits of no 'discontinuities' or 'reversals' ([2002], p. 79). (See Stephens [2004] for a critique of this position). Instead, I favor accounts of causal processes that allow for probabilistic causation; the accounts described above are at a minimum amenable to this sort of interpretation.

the fact that natural selection is statistical (itself an uncontroversial claim) does not mean that it is *purely* statistical. Natural selection can be both statistical and causal, as the following discussion will show.

The montane willow leaf beetle is indigenous to western North America from the Yukon to California, and it can be found on willow shrubs or trees that commonly grow in separated patches of boggy or moist habitats (Rank [1992]). It can fly, but it rarely does so; thus, there is limited migration between localities. Because of the patchiness of the beetle's habitat and the limited migration between patches, there is the potential for adaptation to local microclimates. Thus, in each of the studies of the montane willow leaf beetle that I will discuss here, samples of the beetle have been taken from different locations. In an early study, Nathan Rank collected beetles from three different drainages in the eastern Sierra Nevada mountains: Rock Creek (northernmost), Bishop Creek (central) and Big Pine Creek (southernmost). Using gel electrophoresis, Rank first screened twenty-two enzyme loci and found polymorphism among seven of them. Two were dropped from consideration 'because the bands were not clearly readable for scoring of putative genotypes' (Rank [1992], p. 1099), leaving five readable loci. Then, for each of the five scorable loci, Rank measured the amount of differentiation among the three localities, and found that one locus, phosphoglucose isomerase (PGI), showed differentiation among the three drainages that was ten times greater than the differentiation at the other loci. In addition, there was a striking pattern to the differentiation at the PGI locus. The allele that was the most common in Rock Creek was the least common in Big Pine Creek, and vice versa: the allele that was the least common in Rock Creek was the most common in Big Pine Creek.

Rank states that his statistical analysis 'suggests' that the PGI locus was undergoing natural selection, but his claim for natural selection is no more conclusive than that. Perhaps this is because, as he states at the outset: 'Even in the absence of selection, genetic drift among small-sized populations increases their likelihood of becoming genetically differentiated' (Rank [1992], p. 1097).

I would argue that Rank's suggestion—that the population is undergoing natural selection at the PGI locus—is justified. Why should the genetic differentiation be greater at one locus than at any other locus? One likely explanation seems to be that whereas the alleles at the other loci are drifting, producing relatively low amounts of differentiation, at the PGI locus selection is increasing favored genotypes, depending on the locale.[10] The fact that

[10] Ben Jantzen ([personal communication]) has suggested that selection at the PGI locus is not necessarily the best explanation for the statistical pattern; the loci with little differentiation could be the result of selection (presumably directional selection that has driven the alleles in

the allele that is predominant in the north is the one that is rare in the south (and vice versa) is further support for a hypothesis of natural selection. Perhaps some of the PGI genotypes are better adapted to the northern location and others are better adapted to the southern. Perhaps, as Rank speculates, it has something to do with differing temperatures.

However, the *tentativeness* of his suggestion is also justified. This is because such speculations, no matter how reasonable, are far from conclusive. As Rank states, the results that he observed could be due to drift. In other words, they could be due to an indiscriminate sampling process (Beatty [1984]), a process whereby physical differences between organisms are causally irrelevant to differences in reproductive success (Millstein [2002]). The population that survives to reproduce is always but a subset of the population that is born, and the smaller that subset (or 'sample'), the more likely it is that it is unrepresentative. If by chance this sampling process at the PGI locus has occurred differently in the north than it has in the south, then differentiation would be the result. Rank's statistics suggest selection at the PGI locus, but they could be the result of drift. That is to say, the different processes of selection and drift can produce the same outcome (Millstein [2002], [2005]), the outcome in this case being the same pattern of differentiation among populations.[11]

Of course, observing the populations over a number of generations might provide more definitive statistical evidence for natural selection. However, besides being costly and time-consuming, such studies are not always as illuminating as one might hope. Fluctuations in gene frequencies from one generation to the next might look like the result of drift, yet be the result of selection in a fluctuating environment. (Again, the different processes of selection and drift can produce the same outcome). When the environment fluctuates, genotypes that were previously favored may no longer be favored, and vice versa, producing a fluctuation in gene frequencies over time that mimics the effects of drift. On the other hand, continued increase of a favored genotype might look like the result of natural selection, yet be the less likely (but not improbable) result of drift. Just as a series of fair coin tosses may turn up mostly heads, so may drift produce a 'streak of good luck' for one type over the others (Millstein [2000]).

question to fixation), whereas the more polymorphic locus could result from drift in the absence of selection. This is a fair point. Rank does not elaborate here, but he does cite Slatkin ([1987]) in this context, and Slatkin emphasizes that whether drift leads to differentiation depends on the population size and gene flow (migration). Slatkin argues that for a given population, when you have a number of loci with low differentiation, this is an indication that gene flow is high enough to counteract drift; thus, if there is one locus with significant differentiation, it must be the result of a process that can overcome gene flow (i.e. selection). In any case, if it did turn out that Rank's initial selection hypothesis is not the most likely hypothesis, that lends even more weight to my subsequent point that the statistical analysis alone is insufficient and that one must have an understanding of the causal influences on the population.

[11] I use the terms 'process' and 'outcome' rather than the more common 'process' and 'product' to avoid the teleological implications of the term 'product.'

As biologist Ward Watt has documented, much of the debate concerning the prevalence of selection versus the prevalence of drift (one aspect of the so-called 'neutralist–selectionist' debate) centered over allozymes (forms of an enzyme encoded by different alleles), and indeed, allozymes are the focus of Rank's study, as I will explain below. However, Watt ([1995], p. 873) states: 'It soon became clear that, given feasible sample sizes, even quite different models of allozyme variation could not be resolved with any statistical power.' As a result of this failure, Watt throws down a gauntlet, 'Evolutionary biologists are challenged to bring mechanistic biology into Darwinian context, to move beyond the remains of the ''neutralist–selectionist'' controversy and of purely formal approaches to evolution, to seek biologically based generalities about the evolutionary process' ([1995], p. 869). Watt urges biologists to seek the *causal basis* for claims about selection, and in this way provide a more solid basis than the statistical analysis of population genetics alone can provide. In this context, he discusses Rank's ([1992]) study of the montane leaf beetle in particular:

> 'Rank (1992) ... found substantial local allozyme frequency differentiation, even among adjacent drainages in the same mountain range. Rank's results could be due to local drift, local microhabitat specialization, or a combination, subject to further study—though gene-specific variation in the extent of frequency differences led him to postulate local selection, at least on the PGI gene. In [Rank 1992 and a similar study by Costa and Ross 1994], use of allozymes to probe population structure has ruled out some possibilities, while leaving the *need for further specific work to distinguish among other causal scenarios*—as the authors have been careful to state' (Watt [1995], p. 878; emphasis added).

I mention these quotations from Watt's essay not to make an argument from authority, but to illustrate that my claim that natural selection is a causal process is consistent with at least some biological practice (showing the role that natural selection as a causal process plays in their beliefs, in their methods, and in the conclusions drawn).[12] Indeed, Watt ([1994],

[12] Chris Stephens ([personal communication]) has suggested to me that a less 'naturalistic' philosopher—perhaps even some of the philosophers to whom I am responding in this section—might accept that causality played a role in biological practice and yet not be convinced on the metaphysical point about the role of causality in selection. Here my response is two-fold: (1) WALM make ample use of coin tosses in their arguments. I would think that biological practice is more relevant to conclusions about selection and drift than coin tosses, which are a very poor model for selection. Coin tosses involve only two 'types' (heads or tails)—a point whose relevance I will demonstrate below—and there is no reproduction involved; moreover, the very concept of 'selecting' is virtually absent. The idea of selecting balls from an urn that biologists use, where the balls can vary in their color or other attributes, is also imperfect (it too lacks a reproductive aspect), but it is preferable. (2) Although the appeal to biological practice is part of my argument in this section, it is not intended to be the whole argument. As I discuss in subsequent paragraphs, statistics alone fail to capture important facts about the biological population.

[1995]) documents a number of recent studies that probe the causal (or 'mechanistic') basis for claims concerning natural selection, and an even greater number of such causal studies are documented in John Gillespie's ([1991]) *The Causes of Molecular Evolution*. Many biologists see such studies as necessary, and for good reason. Statistical analysis alone (an analysis of outcomes) is usually inadequate to answer the questions posed.

In fact, in subsequent studies Rank and his colleague Elizabeth Dahlhoff *have* sought out a causal basis for the distribution of beetles between the three drainage areas.[13] Rank and Dahlhoff do, of course, collect more statistical data, and they find that the distribution pattern of PGI genotypes that was present in the earlier study (performed in 1988) is similar to the distribution of PGI genotypes in a later study (performed in 1996). In fact, there were increases in the previously favored genotypes, further strengthening the claim for natural selection (Dahlhoff and Rank [2000]; Rank and Dahlhoff [2002]). However, most of their recent efforts have gone towards establishing the *causal* basis of selection at the PGI locus. Rock Creek, the northernmost drainage, is at a higher elevation and cooler than the other drainages, whereas Big Pine Creek, the southernmost drainage, is at a lower elevation and warmer than the other drainages. PGI 1-1, a homozygote, is common at the former, but rare at the latter; PGI 4-4, another homozygote, is common at the latter but rare at the former. Bishop Creek is intermediate between the two in terms of elevation and average temperatures, and there, a third PGI genotype predominates: the heterozygote PGI 1-4 (followed by PGI 1-1, then PGI 4-4).

These differences in genotype distribution and temperature are paralleled by differences in the amounts of heat shock protein produced by the different PGI genotypes. Heat shock proteins protect other proteins from heat-related damage 'by refolding partially unfolded proteins into their functional state' (Dahlhoff and Rank [2000], p. 10056), but they also appear to enhance tolerance to cold temperatures (Rank and Dahlhoff [2002]). Rank and Dahlhoff's laboratory experiments show that the homozygous PGI 1-1 genotype expresses heat shock proteins at a lower temperature (reaching a peak of heat shock protein expression at 30°C) than the homozygous PGI 4-4 genotype (reaching a peak of heat shock protein expression at 36°C), with the heterozygous PGI 1-4 genotype expressing heat shock proteins at intermediate temperatures (reaching a peak of heat shock protein expression at 33°C) (Rank and Dahlhoff [2002]). Furthermore, when they exposed the beetles to extremely cold temperatures in the laboratory, most of the female PGI

[13] According to Nathan Rank ([personal communication]), he and Elizabeth Dahlhoff see themselves as part of a research program whose approach is similar to that described by Feder and Watt ([1992]). This approach looks at the functional and physiological consequences of genetic variation and tracks them upwards to organisms living under natural conditions in order to see how physiological differences play out in differences in reproductive success.

1-1 and 1-4 genotypes survived, but half of the female PGI 4-4 genotypes died (Rank and Dahlhoff [2002]).

These causal factors (the differing abilities of the different genotypes to survive hot and cold temperatures) are consistent with the statistical data. That is, PGI 1-1 has the greatest ability of the three genotypes to survive cold temperatures, and it is the genotype that is the most prevalent in the coldest region; PGI 4-4 has the greatest ability to survive warmer temperatures, and it is the most prevalent in the warmest region, and the heterozygote PGI 1-4 that does best at intermediate temperatures is indeed the most prevalent in the intermediate region. It is important to emphasize that the differing capabilities of the different beetle genotypes to survive differing temperatures were determined in the laboratory, thus providing an independent line of evidence. Moreover, as Rank and Dahlhoff ([2002], p. 2285) note, the laboratory results 'suggest that differences among *Pgi* genotypes in thermal physiology are responsible for *Pgi* allele frequency change in nature.' Thus, the causal data provide support for the claim that these populations are undergoing natural selection by supplying a causal explanation for the observed statistical pattern, lending credence to the claim that the pattern was not produced by some other means.

Rank and Dahlhoff's studies are extensive and well done. They include at least three of five components of Brandon's ([1990], p. 165) 'ideally complete adaptation explanations': (1) *evidence that selection has acted* on PGI genotypes (the statistical evidence, taken over a time period); (2) *an ecological explanation* of why some PGI genotypes are better adapted than others in different environments (they have identified, using Brandon's terms, the *selective agent* as the temperature and the *adaptation* as the 'ability to withstand heat or cold'); and (3) evidence that the ability to withstand heat or cold is *heritable*.[14] For what it is worth, Rank and Dahlhoff also discuss the underlying mechanism of heat and cold tolerance in the beetles; Brandon does 'not consider [a physiological or biomechanical explanation of mechanisms] a necessary part of the evolutionary explanation of the adaptation' (Brandon [1990], p. 166). However, it seems clear that the ecological explanation needs to be enhanced; otherwise, how are we to be confident that we have correctly identified the 'selective agent' and the 'adaptation'? Perhaps for example, it only *appears* as if there is selection at the PGI locus, when in

[14] The fourth component is 'information about the structure of the population from both a genetic and a selective point of view, that is, information about patterns of gene flow and patterns of selective environment' and the fifth component is 'phylogenetic information concerning what has evolved from what' (Brandon [1990], p. 165). With regard to the fourth component, Rank and Dahlhoff indicate that the montane willow leaf beetle can be found in separated patches of boggy or moist habitats, with limited migration between patches. This goes at least part of the way towards satisfying the fourth component. They do not, to my knowledge, provide any phylogenetic information about the beetle.

reality selection is occurring at another, linked locus. We need either to perform a laboratory experiment to demonstrate that the selective agent indeed acts on the phenotype in the way we think it does, or we need to provide the underlying mechanism to show that the genotypes have the abilities that we say that they do, or both. Rank and Dahlhoff do both; the laboratory experiment is described above, and they explain that the different PGI genotypes' differing abilities to withstand heat and cold are the result of the production of differing amounts of heat shock protein at different temperatures. This causal and mechanistic information, together with the other information presented, provides strong evidence for Rank and Dahlhoff's conclusions.

However, it should be noted that Rank and Dahlhoff's research is ongoing, and it appears to be complicated by the fact that the beetles acclimatize to different temperatures and by differing responses from males and females of different ages. Nonetheless, two points emerge clearly. First, finding a causal mechanism is a crucial piece of the natural selection story; the statistical results are not sufficient on their own. Second, there seems to be a clear connection between the differing temperatures, the amount of heat shock protein produced by the different genotypes, and the subsequent survival ability of different genotypes. In other words, beetles that have an enhanced ability to survive cold temperatures (via heat shock proteins) are selectively favored in the northernmost region, whereas beetles that have an enhanced ability to survive warm temperatures are selectively favored in the southernmost region.

The causal story is essential to understanding the statistical pattern. Moreover, if there were no such causal differences present—if there were no differences in survival or reproductive ability between the different genotypes—*selection could not occur*.[15] Instead, the observed patterns would be attributable to drift or some other evolutionary process. It is for this reason that natural selection should be characterized as a discriminate sampling process whereby physical differences between organisms are causally relevant to differences in reproductive success. Drift, by contrast, is an indiscriminate sampling process whereby physical differences between organisms are causally irrelevant to differences in reproductive success (Millstein [2002], [2005]).

4 The montane willow leaf beetle: a population-level story

Even if I am correct that natural selection is a causal statistical process, rather than a *purely* statistical one, the question remains as to whether causality acts at the population level or at the individual level.

[15] Of course, there may be differences in ability that we are unable to detect, but that is a different issue.

There are conflicting intuitions about this question. On the one hand, it might seem as though natural selection is about individuals: their lives, their deaths, their reproductive successes or failures. On the other hand, an oft-quoted, if not universally accepted passage from Sterelny and Kitcher ([1988], p. 345) suggests otherwise:

> In principle, we could relate the biography of each organism in the population, explaining in full detail how it developed, reproduced, and survived, just as we could track the motion of each molecule of a sample of gas. But evolutionary theory, like statistical mechanics, has no use for such a fine grain of description: the aim is to make clear the central tendencies in the history of evolving populations.

The first intuition runs into some immediate problems. Although it is true that these individual-level causal events (living, dying and reproducing) form the basis for natural selection, they do not constitute natural selection itself. Certainly, they do not constitute *evolution* by natural selection. Definitions of evolution differ, but one common definition (at least among population geneticists) is 'change in gene frequencies from one generation to the next'. 'Gene frequency' is a property of a population—it refers to the percentage of individuals of each type within a particular population. Moreover, although an individual organism's properties may change over the course of its lifetime—the organism may grow, change color, become stronger, even lose a limb—it cannot be said to evolve. It is only *populations* that can evolve through a change in genetic composition from one generation to the next.

The second intuition, however, is not unproblematic either. I agree with the spirit of Sterelny and Kitcher's quote, assuming 'tendencies' is meant in a causal sense, but its wording is somewhat unfortunate. As Rosenberg ([1994], p. 62) notes, it is theorists who have aims, not theories, which makes the population-level character of evolutionary theory sound purely instrumental. However, as Rosenberg also notes, it is not clear that Sterelny and Kitcher intend any such instrumentalism. Furthermore, the same essential point can be made by examining evolutionary concepts and processes themselves, in addition to biological practice, as I will do in this section.

I imagine that BR would remain unmoved by my concerns with the first intuition; they have argued that we should 'treat selection as a contingent causal process in which individual fitness differences are the causes and subsequent population differences are the effects' ([2004], p. 710). To show that this claim does not hold up, I will again look to Rank and Dahlhoff's studies of the montane willow leaf beetle. As we have seen, these studies focus on the causes inherent in selection—but are the selective processes acting on *individual* beetles or acting on *populations* of beetles?

In order to answer this question, we must first characterize what it means to say that selection acts on individuals. However, there are at least two

possible ways to do this, which I will term 'naïve individualism' and 'sophisticated individualism'. BR, it should be noted, endorse the sophisticated individualist position; however, an examination of the naïve position (one I think is implicit in some accounts of natural selection) will highlight the importance of what I call *the comparative nature of natural selection*. I will take up naïve individualism in Section 4.1 and sophisticated individualism in Section 4.2.

4.1 Response to 'naïve individualism'

On the naïve individualism view, it might be granted that, as a matter of practice, evolutionary biologists focus on populations and characteristics of populations. However, the naïve individualist asks whether biologists do this because that is where the real causal story is, or because it is easier and more manageable. Ultimately, the naïve individualist believes that natural selection is about the lives, deaths and reproductive successes of individual organisms. However, the naïve individualist points out that it would be overwhelming, time-consuming and expensive for biologists to try to observe, understand and document each individual's death or reproductive success. On this view, then, biologists look at the general characteristics of populations purely for practical reasons; their choice has nothing to do with the process of natural selection itself. In principle, we could follow the life history of organisms and know what caused them to survive long enough to reproduce or die before reproducing, just as we could (in principle) study the lungs of smokers throughout their lifetimes in order to determine whether smoking was the cause of their lung cancer.[16]

However, the naïve individualist view does not succeed *even given individualistic assumptions* because it falsely assumes that the natural selection case is similar to the smoking case. Whereas we could (in principle) monitor an individual's lungs to determine whether smoking was the cause of her lung cancer, the best we could do (again, in principle) in the realm of selective scenarios is to say whether an individual organism's heritable physical characteristics play a causal role in its survival and reproductive success. But this information would not allow the individualist to determine whether the organism was selectively favored, selectively disfavored, or subject to selection at all. An organism's heritable physical characteristics might cause it to survive long enough to reproduce, but if a second organism's different heritable physical characteristics cause it to have *greater* reproductive success than the first organism, the first organism will not be selectively favored.

[16] The position represented by this counter-argument might seem to be that of a determinist; however, an indeterminist might take such a position as well, maintaining that we could follow the causal influences on the organism even though those causes would not fully determine the outcome.

Similarly, if organisms' different physical characteristics each cause reproductive success in the given environment such that their reproductive successes are equal, neither of the organisms will be selectively favored over the other. (Of course, the populationist would not grant that an individual organism could be selectively favored over another individual organism; the point here is that naïve individualism yields incorrect results even given individualist assumptions).

What this discussion shows is that you cannot follow the life history of *one* organism and determine whether or not it has been selectively favored. That determination requires *comparison* with the other organisms in the population, whereas determining whether one person's smoking caused her lung cancer does not require comparison. Selection is a game where what matters is how well you are able to do relative to others in the population, not how well you are able to do in some absolute sense. To elaborate on this point, and to show how it supports a population-level account, let us consider three possible scenarios.

In the first scenario, a population consists of individuals all of the same genotype. In this scenario, there would be no selection, because variation is a necessary condition for selection, as has been acknowledged in every definition of selection from Darwin to the present day. Indeed, without variation—without differences—there is nothing to select among.[17] Thus, in his initial study, Rank ([1992]) screened for polymorphism at twenty-two enzyme loci. However, only seven of those were polymorphic. The remaining fifteen were not studied any further; they could not exhibit selection because there was no variation for selection to act upon. Note that 'variation' is a property of a population, not an individual, manifested by organisms differing in their heritable physical traits.

In the second scenario, a population consists of individuals of different genotypes, yet the different genotypes are equal in their abilities to survive and reproduce. There will be no selection in this scenario, either; again, it is universally acknowledged that selection requires differences in abilities to survive and reproduce (what some would term 'fitness differences'—I avoid the term to prevent entanglement in disputes over the definition of fitness), and yet under this scenario, there are no such differences for selection to act on. A lack of significant differences in genotype frequencies in a population

[17] In this first scenario—indeed, throughout this essay—I am assuming that selection requires *heritable* variation, a point on which there is a lack of universal agreement. (It is my belief that selection without heritability is trivial—choosing which pen to use becomes an instance of selection—but this is a digression that I will avoid here). Someone who does not hold that selection requires heritable variation might argue that although the individuals are identical genotypically, they differ phenotypically, and thus there could be selection among the different genotypes. However, note even on this view, variation is still a necessary condition; if there were no phenotypic differences then there could be no selection.

can serve as evidence (although not definitive evidence) for equal abilities to survive and reproduce. Thus, in the montane willow leaf beetle studies, Rank, and subsequently Rank and Dahlhoff, rejected four of the five scorable polymorphic loci as obvious candidates for selection—there was variation at these loci, but there were no significant differences in frequencies of the different genotypes and thus, no obvious benefits conferred by being one genotype over another (although, of course, it is possible that as a result of other factors such differences simply failed to manifest themselves, or that the population had undergone selection to fixation in the past). Scenarios of this second type illustrate that *it does not matter if a particular individual has a superb ability to survive and reproduce—there will be no selection at all if other individuals who are physically different have the same ability*. In other words, it is one genotype's ability to survive and reproduce relative to other genotypes within the population that leads to selection.

In the third scenario, there are different genotypes, and the different genotypes differ in their survival and reproductive abilities. In this scenario, selection *can* occur, acting on the different abilities in the population. Differences in genotype frequencies can suggest (non-definitively) that selection has occurred; this was the case with the PGI locus in the beetle studies, and thus this was the locus that Rank and Dahlhoff chose to study in order to determine if selection was in fact at work. That is, it was the locus that was the most promising, the one whose polymorphism was most likely to have been the result of differing abilities. Indeed, as discussed above, Rank and Dahlhoff sought to establish these differing abilities by exposing the genotypes to different temperatures in the laboratory and measuring the production of heat shock proteins. Had the different genotypes not produced differing amounts of heat shock protein, the researchers would have been forced to look for other differences in abilities among the genotypes or, if they failed to find any such, to conclude that there were no differences in abilities among the genotypes (i.e. they would be forced to conclude that the second scenario had obtained and that selection was not responsible for the observed variation). But differences in the production of heat shock proteins *were* found (i.e. the third scenario obtained), providing the necessary conditions for selection. Thus, in scenarios of this third type, there is the possibility for selection, but only because of the ability of some genotypes as compared with the ability of other genotypes within the population.

These three scenarios taken together describe the comparative nature of natural selection, and its comparative nature entails that it is a population-level process. Natural selection requires that some genotypes have a greater ability to survive and reproduce relative to other genotypes. Otherwise, there

is nothing for 'nature' to 'select', or there is no basis on which 'nature' is able to 'select.' An individual organism might have heritable traits that enable it to survive and reproduce in its environment, but selection can occur only if there are differences in abilities in the rest of the population. On the other hand, 'variation in genotypes' is a property of the population, as is 'variation in the abilities of genotypes.' These population-level properties are the causal engine of selection. This is causality at the population level; the differing physical abilities, an attribute of the population, do the causal work of selection.[18]

In terms of the beetle case, we have selection when the differing production of heat shock proteins by the different genotypes (yielding differing abilities to withstand hot and cold temperatures) is causally relevant to differences in reproductive success among the genotypes. Or, more generally, we have selection when heritable differences in physical characteristics (yielding differences in survival and/or reproductive abilities) are causally relevant to differences in reproductive success. In short, I have described a discriminate sampling process (Millstein [2002]).

However, Matthen and Ariew seem to imply that sampling processes are at the individual-level; Matthen and Ariew ([2002], p. 62) state: 'And so one might think that one can fix the exact role of vernacular fitness and of drift by looking more minutely at individual events and determining when discriminate sampling has been at work, and when indiscriminate.' But this view is mistaken; 'sampling' implies a causal process that acts on a population.[19] More specifically, it is the process that acts on a population as a whole that determines whether the sampling is discriminate or not.

To see why sampling is population level, consider, for example, a redwood tree that is killed as a result of a fire in which many, but not all, of the redwood trees in the forest were killed. Consider two possible scenarios under which this could occur: (1) there is a heritable physical characteristic of the tree in question that caused it to be killed; this characteristic is shared by most of the other trees that were killed (and it caused their deaths as well); (2) there is no heritable physical characteristic shared by the tree in question and most of the other trees that were killed that is causally relevant to their death by forest fire. The first is discriminate sampling; the second is

[18] Or, they do at least some of the causal work; below, I will describe a case that illustrates the causal action of the environment on a population.

[19] It might seem odd that I would find reason to disagree with Matthen and Ariew ([2002]) in this section of the paper, given that, as I noted above, we agree that natural selection is a population-level affair. But their characterization of sampling—a causal process—as individual-level rather than population-level enables them to overlook the fact that natural selection is a population-level *causal* process.

indiscriminate sampling. And yet the causal process that acted on the individual tree in question is the same in both scenarios: death by fire. There is no way we could determine, by following the life history of the tree alone, whether it was killed by a discriminate sampling process or an indiscriminate sampling process. But the point is more than epistemic; there simply *is* no sampling process acting on the tree alone. Rather, there is a forest fire that acts on the redwood population, either killing discriminately or killing indiscriminately. As Shapiro and Sober ([forthcoming]) emphasize, 'Sometimes [two individuals] x and y experience the same token selection process because they causally interact; at other times they participate in the same selection process because they are affected by a common token cause.' Shapiro and Sober rightly note that this is implicit in Darwin's idea of the 'struggle for existence', which encapsulates both direct competition between individuals[20] as well as cases such as a plant struggling on the edge of a desert. With regard to the latter, Shapiro and Sober suggest that it is 'entirely natural' to point to the drought as a single process affecting individual plants. My suggestion here is that it is likewise 'entirely natural' to see the forest fire as a single process acting on the trees; together with the fact that the action of that single process acts discriminately or indiscriminately only in the context of the population as a whole (here again, we see the importance of the comparative aspect of selection), it would seem that discriminate sampling, i.e. selection, is a population-level process.

4.2 Response to 'sophisticated individualism'

However, my account of natural selection as fundamentally comparative raises the possibility of a more sophisticated individualist counter-argument to my claim that natural selection is a population-level process. Perhaps what needs to be done is not to follow the life history of individual organisms (as the naïve individualist would have us do), but to perform *pairwise comparisons* between individual organisms. BR espouse this view in a recent paper. According to BR, organisms are compared two at a time in terms of their relative abilities to solve 'design problems' in a given environment. Thus, their account is causal, takes into consideration the comparative nature of natural selection, and yet preserves the individual-level account of selection. Or, at least it purports to—it seems to me that their view collapses into the population-level account, as I will show.

[20] Indeed, where selection does involve such direct competition—whether it be males engaged in a physical competition for females, or organisms competing for scarce resources—it becomes even harder to make the case for individual-level selection.

How would these pairwise comparisons proceed? BR do not tell us exactly how to do this, so I will try to extrapolate from the assumptions of their account. BR state that we need to consider 'ecological fitness', where 'a is fitter than b in [environment] E means that 'a's traits result in its solving the design problems set by E more fully than b's traits' (Bouchard and Rosenberg [2004], p. 699). And because we are comparing individuals, we would need to compare each individual in the population, 'taken one at a time', (Bouchard and Rosenberg [2004], p. 709) to every other individual in the population. For example, if a population had three individuals, *A*, *B* and *C*, we might begin by determining that *A* is better able to solve design problems in the given environment than *B* (has greater ecological fitness), so we expect that *A* will have more offspring than *B*. Next, we might determine that *B* has greater ecological fitness than *C* (again, by comparing their abilities to solve design problems), so we expect *B* to have more offspring than *C*.

Finally, we must compare *A* with *C*. In 'transitive' populations (a population where if *A* is fitter than *B* and *B* is fitter than *C*, *A* will also be fitter than *C*) we can say that *A* is fitter than *C* in *this* environment; however, if *A* solves design problems through a different mechanism than *C*, *A* may not be fitter than *C* in a future environment. Thus, in order to be able to make predictions under conditions where the environment is changing, all individuals must be compared with all other individuals. Given that BR intend 'ecological fitness' to be a qualitative, causal determinant, and not just a quantitative 'head count' ([2004], p. 709), the latter comparison seems to be a necessary step.

Indeed, the existence of 'non-transitive' populations highlights the absolute necessity of comparing all individuals. Non-transitive populations can be compared with the children's game of 'Rock, Paper, Scissors' where rock crushes scissors, scissors cut paper, but paper covers rock (Kerr et al. [2002]). In other words, none of the three is superior to the other two; each is superior to one, but inferior to another. Kerr et al. ([2002]) study a population of *Escherichia coli* that manifests this non-transitivity. The population contains three types of cells: (1) colicinogenic cells (C), which are cells that can produce a toxin, colicin; (2) sensitive cells (S), which are cells that can be killed by the colicin; and (3) resistant cells (R), which are immune to the colicin. Under certain conditions, C is superior to S (the colicin released by the C cells kills the S cells), but S is superior to R (because the mechanism that provides R with immunity from the colicin hampers its nutrient uptake), while R can displace C (because the mechanism that allows C to release its toxin incurs a cost that hampers its growth rate). When C, S and R are located in 'clumps', or patches, this polymorphism can be maintained (however, in a 'well-mixed' system, C drives S extinct and then R

outcompetes C) (Kerr et al. [2002]). If one were only to compare, for example, C to S and S to R, one would miss the relationship between C and R.[21]

Thus, in general, for populations of N individuals, you would need to make $(N)(N-1)/2$ comparisons, that is, enough comparisons so that each individual is compared with every other individual in the population. As a practical matter, this would never be done, but in principle, so the argument goes, it *could* be done.

After all the pairwise comparisons are completed, we must now determine which individuals will be selectively favored and which individuals will be selectively disfavored. I cannot see any way of doing this other than poring over the pairwise comparisons to establish a *complete* ranking of individuals, as far as that can be done. That is, we would need to determine which individuals are the best at solving design problems, which are the worst, and which are intermediate. The reason for this is that, even though A may be ranked higher than B, both A and B may be worse at solving design problems as compared with other individuals in the population, and thus be selectively disfavored. Conversely, *both* may be selectively favored if they are better at solving design problems as compared with other individuals in the population. We can imagine a huge ranking of all the individuals in the population, with some ranked higher than others and some ranked equally, and again, we would expect reproductive successes or failures accordingly.

Two points need to be emphasized here. First, philosophers often speak as though there are only two genotypes in the population (or at best, three—two homozygotes and one heterozygote—but then it often turns out that we have only two phenotypes to consider). Perhaps they are following the simple models that can be found in the early chapters of introductory population genetics textbooks, but we have to remember that biologists use these simple models to make the mathematics more tractable, not as a reflection of biological reality. Skip to a later chapter, and one finds comments such as the following:

> There are often more than two alleles at one locus. Indeed, the recent surveys of enzyme polymorphism ... reveal many loci with 4 to 5, even 10 alleles (Roughgarden [1996], p. 101).

Indeed, in Rank and Dahlhoff's study, there are more than two genotypes; Dahlhoff and Rank identified eight genotypes at the PGI locus ([2000]), and the three genotypes that they focused on each had a different phenotype (a different ability to withstand heat and cold). Now, to be fair, I think that most philosophers recognize that the two-allele, one-locus model is a

[21] Thanks to Patrick Forber for suggesting that the essay by Kerr et al. is relevant to my arguments.

simplification; what they fail to recognize is that the simplification causes them to overlook the comparative nature of selection and its implications.

Second, the careful reader will note that throughout this paper I have used the terms 'selectively favored' and 'selectively disfavored' rather than the more common 'selected for' and 'selected against.' Once the comparative nature of selection is recognized along with the fact that we are, more often than not, dealing with multiple genotypes, it becomes obvious that selection is not a binary affair of 'selected for' or 'not selected.' Rather, some genotypes may be favored *as compared with other genotypes*; but whether those genotypes are ultimately favored depends on the composition of the population at large (again, because a third, fourth or fifth genotype may be more successful than either of the first two).[22]

Perhaps I have misunderstood how the pairwise comparisons are supposed to work, but it would seem that this collection of individuals, with some individuals ranked higher than other individuals, is a population. If in the end, the real selective story is known only when all the pairwise comparisons have been done and the ranking completed, then any given pairwise comparison is superfluous—it is only an intermediate step that leads to the real selective story at the population level. Or, to put the point another way, if you do not establish the rankings, then you do not know which individuals are selectively favored. But once you do establish the rankings, you are no longer dealing with individuals; you are dealing with a population.[23]

Moreover, of those individuals who are ranked equally, some will solve design problems in the same way, and some may solve design problems in a different way (but have the same ability to survive and reproduce). As noted above, it would be important to distinguish these two groups because if the environment changes, these groups may no longer have the same ability to survive and reproduce. In other words, it would be crucial to know the different *genotypes* in the population, and not just which individuals had equal abilities in the given environment. Thus, the rankings obtained should be

[22] Perhaps an even better alternative is to distinguish between genotypes (or genes) that increase in frequency and genotypes (or genes) that decrease in frequency; Hartl and Clark ([1989], p. 169) note that 'an allele will increase in frequency provided its marginal fitness exceeds the mean fitness of a population.' Geometric mean fitness is, of course, a characteristic of the population rather than of any given individual; this too is evidence that selection operates at the population-level.

[23] Furthermore, as Elliott Sober has suggested ([personal communication]), if you compared the fitnesses of all individuals with each other, you would thereby discover the fitness ordering of different trait combinations, but not necessarily of single traits. The approach of ranking genotypes, on the other hand, allows one to isolate the differences made by a single trait, if any.

rankings of the genotypes in the population, and indeed, this is what Rank and Dahlhoff provide us, for each of the microclimates under study. The end result seems to be that pairwise comparisons end up dealing with populations and population parameters (but, of course, the long way around; even the 'Rock, Paper, Scissors' study was not done by comparing every individual with every other individual). The individual-level account collapses into the population-level account.[24] The conclusion to be drawn is that natural selection is a causal process that operates at the level of populations.

5 Conclusion

In this paper, I have argued that natural selection must be understood as a causal process operating at the population level. First, even though population-level causation may be controversial in some circles, it is no more problematic for selection in general than it is for frequency-dependent selection in particular, a widely accepted and well-documented biological phenomenon. Moreover, population-level causation in natural selection can be understood in terms of a number of popular contemporary accounts of causation. Second, statistics alone are insufficient to capture natural selection, as my discussion of Rank and Dahlhoff's studies of the montane willow leaf beetle showed; the initial statistical results were compatible with either natural selection or random drift. Thus, as a theoretical matter, causality is a necessary component of natural selection; furthermore, good biological practice seeks out the causal basis for selection claims. Third, causation in natural selection operates at the population level, as the beetle studies also illustrate, because of the inherently comparative nature of the selection process. Again, there is no way to make sense of selecting for an individual if selection is determined only by what is happening in the population as a whole. Attempts at capturing selection as an individual-level causal process collapse into a population-level account.

Of course, none of this is to deny that populations are composed of individuals or that causes are acting on those individuals. Nor is it to deny that individual-level causal events (living, dying and reproducing) are in some sense responsible for selection. What I deny is that such individual-level causal events constitute the selection process itself. A long beak might be causally relevant to the survival (and ultimately, reproduction) of a particular finch because it allows the finch to feed on a particular type of seed. However, in

[24] There is a further concern with BR's account. In order to distinguish selection from drift, they suggest that with selection, there is a random distribution of initial conditions, and that drift comes into play when this distribution is non-random. However, distribution of initial conditions is a property that occurs across a population, not a property of an individual. Thus, once again BR's account collapses into a population-level account.

one population, that type of organism might be selectively favored, whereas in a different population, that type would be selectively disfavored, and in a third, it might not be selected at all. Furthermore, we can sum up the number of finches who survived because of their long beaks and the number of finches with shorter beaks who perished without reproducing (or who produced fewer offspring). But this is just to assert that some individuals live and reproduce prodigiously and some do not. If that had been all that there was to Darwin's theory of natural selection, not only would it not have been controversial, it would not have been given any attention at all, because it would have expressed a triviality. One of the things that makes the theory non-trivial is the invocation of its causal basis; a causal basis that, properly understood, does not comprise the causes acting on individuals, but instead comprises the causes that act on the population as a whole: the relative physical differences between genotypes in the population that yield differing relative abilities to survive and reproduce. And that is just to reassert what I have argued throughout: natural selection is neither a *purely* statistical (acausal) population-level summation, nor is it a process of individual-level causation. Natural selection is, properly understood, a process that exhibits population-level causation.

Acknowledgements

I would like to thank the following people for their helpful comments and discussion regarding various versions of this paper: Nathan Rank, Elizabeth Dahlhoff, Elliott Sober, Bruce Glymour, Marcel Weber, Michael Dietrich, Rob Skipper, Stuart Glennan, Frédéric Bouchard, Ben Jantzen, Patrick Forber, Chris Stephens, Anya Plutynski and members of the audience at ISHPSSB 2003, held in Vienna, Austria. Special thanks go to a sharp-eyed undergraduate student, Duane Meehan, for suggesting that I might be interested in Rank and Dahlhoff's studies of the montane willow leaf beetle. Revisions to this paper were supported by a Faculty Support Grant from the California State University East Bay.

Department of Philosophy
University of California, Davis
One Shields Avenue
Davis, CA 95616, USA
RLMillstein@UCDavis.edu

References

Beatty, J. [1984]: 'Chance and Natural Selection', *Philosophy of Science*, **51**, pp. 183–211.

Bouchard, F. and Rosenberg, A. [2004]: 'Fitness, Probability, and the Principles of Natural Selection', *British Journal for the Philosophy of Science*, **55**, pp. 693–712.

Dahlhoff, E. P. and Rank, N. E. [2000]: 'Functional and Physiological Consequences of Genetic Variation at Phosphoglucose Isomerase: Heat Shock Protein Expression Is Related to Enzyme Genotype in a Montane Beetle', *Proceedings of the National Academy of Sciences USA*, **97**, pp. 10056–61.

Dupré, J. [1984]: 'Probabilistic Causality Emancipated', *Midwest Studies in Philosophy*, **9**, pp. 169–75.

Feder, M. E. and Watt, W. B. [1992]: 'Functional Biology of Adaptation', in R. J. Berry, T. J. Crawford and G. M. Hewitt (*eds*), 1992, *Genes in Ecology*, Oxford: Blackwell Scientific Publications, pp. 365–92.

Futuyma, D. J. [1986]: *Evolutionary Biology*, Sunderland, MA: Sinauer Associates.

Giere, R. N. [1984]: *Understanding Scientific Reasoning*, New York: Holt, Rinehart, and Winston.

Gillespie, J. H. [1991]: *The Causes of Molecular Evolution*, New York: Oxford University Press.

Hartl, D. L. and Clark, A. G. [1989]: *Principles of Population Genetics*, Sunderland, MA: Sinauer Associates.

Hitchcock, C. R. [1996]: 'The Mechanist and the Snail', *Philosophical Studies*, **84**, pp. 91–105.

Hodge, M. J. S. [1987]: 'Natural Selection as a Causal, Empirical, and Probabilistic Theory', in L. Krüger (*ed.*), 1987, *The Probabilistic Revolution*, Cambridge, MA: MIT Press, pp. 233–70.

Kerr, B., Riley, M. A., Feldman, M. W. and Bohannan, B. J. M. [2002]: 'Local Dispersal Promotes Biodiversity in a Real-Life Game of Rock-Paper-Scissors', *Nature*, **418**, pp. 171–74.

Lewis, D. [1973]: 'Causation', *Journal of Philosophy*, **70**, pp. 556–67.

Matthen, M. and Ariew, A. [2002]: 'Two Ways of Thinking About Fitness and Natural Selection', *The Journal of Philosophy*, **99**, pp. 55–83.

Millstein, R. L. [2000]: 'Chance and Macroevolution', *Philosophy of Science*, **67**, pp. 603–24.

Millstein, R. L. [2002]: 'Are Random Drift and Natural Selection Conceptually Distinct?' *Biology and Philosophy*, **17**, pp. 33–53.

Millstein, R. L. [2005]: 'Selection vs. Drift: A Response to Brandon's Reply', *Biology and Philosophy*, **20**, pp. 171–5.

Rank, N. E. [1992]: 'A Hierarchical Analysis of Genetic Differentiation in a Montane Leaf Beetle *Chrysomela aeneicollis* (Coleoptera: Chrysomelidae)', *Evolution*, **46**, pp. 1097–111.

Rank, N. E. and Dahlhoff, E. P. [2002]: 'Allele Frequency Shifts in Response to Climate Change and Physiological Consequences of Allozyme Variation in a Montane Insect', *Evolution*, **56**, pp. 2278–89.

Reisman, K. and Forber, P. [forthcoming]: 'Manipulation and the Causes of Evolution', *Philosophy of Science*.

Rosenberg, A. [1994]: *Instrumental Biology or the Disunity of Science*, Chicago: University of Chicago Press.

Rosenberg, A. and Bouchard, F. [2005]: 'Matthen and Ariew's Obituary to Fitness: Reports of Its Death Have Been Greatly Exaggerated', *Biology and Philosophy*, **20**, pp. 343–53.

Roughgarden, J. [1996]: *Theory of Population Genetics and Evolutionary Ecology: An Introduction*, Upper Saddle River, NJ: Prentice Hall. Third edition; first published 1979, New York: Macmillan Publishing Company.

Slatkin, M. [1987]: 'Gene Flow and the Geographic Structure of Natural Populations', *Science*, **236**, pp. 787–92.

Shapiro, L. and Sober, E. [forthcoming]: 'Epiphenomenalism–The Do's and the Don'ts, in G. Wolters and P. Machamer (*eds*.), *Studies in Causality: Historical and Contemporary*, Pittsburgh: University of Pittsburgh Press.

Sober, E. [1984]: *The Nature of Selection*, Cambridge, MA: MIT Press.

Stephens, C. [2004]: 'Selection, Drift, and the ''Forces'' of Evolution', *Philosophy of Science*, **71**, pp. 550–70.

Sterelny, K. and Kitcher, P. [1988]: 'The Return of the Gene', *The Journal of Philosophy*, **85**, pp. 339–61.

Walsh, D. M., Lewens, T. and Ariew, A. [2002]: 'The Trials of Life: Natural Selection and Random Drift', *Philosophy of Science*, **69**, pp. 452–73.

Watt, W. B. [1994]: 'Allozymes in Evolutionary Genetics: Self-Imposed Burden or Extraordinary Tool?' *Genetics*, **136**, pp. 11–6.

Watt, W. B. [1995]: 'Allozymes in Evolutionary Genetics: Beyond the Twin Pitfalls of ''Neutralism'' and ''Selectionism'' ', *Revue Suisse de Zoologie*, **102**, pp. 869–82.

Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*, New York: Oxford University Press.