Cross-Unit Causation and the Identity of Groups

Bruce Glymour*†

In this article I explore some statistical difficulties confronting going conceptions of 'group' as understood in accounts of group selection. Most such theories require real groups but define the reality of groups in ways that make it impossible to test for their reality. There are alternatives, but they either require or invite a nominalism about groups that many theorists abjure.

1. Introduction. In this article I recount some empirical and associated statistical difficulties confronting going accounts of group selection in respect of the conceptions of 'group' employed by them. In brief, most such theories require real groups but define the reality of groups in ways that make it impossible to test for the reality of the groups employed by a population model. If those definitions are to be taken seriously, no group selection model can ever be employed on real observational data in any reliable fashion; even the diagnosis of the presence or absence of group selection is statistical nonsense given these accounts. There are alternatives, but they require a nominalism about groups that most theorists (although rather fewer practitioners) abjure.

Models of group selection require groups: some mapping of individuals to collections of individuals. On some theories of group selection, those collections comprise merely nominal groups; that is, any mapping will do, so long as every member of the population is mapped to some collection. Consider, for example, models employing neighborhood variables. Each unit,

each oak tree on a hillside perhaps, is mapped to the collection of oak trees that are within, say, 10 meters of the focal unit. Each oak tree has its own neighborhood, so every member of the population is mapped to some (possibly empty) collection. But no oak tree is in its own neighborhood, and any given oak tree might be in the neighborhood of several others. Further, the neighborhoods—the collections of oak trees—need not themselves be in any interesting sense biologically united, need not comprise a 'real' group, apart from the fact that they happen to be all the oak trees within 10 meters of some other oak tree. The group is formed by appeal to a mere Cambridge property, as it were.

Some (e.g., Glymour and French 2009) are content with neighborhoods or other merely nominal groups, and I will call these positions 'nominalist'. But I do not mean by nominalism to require that groups are not real in any biological sense; nominalism merely says that reality, and so reality in any particular sense, is irrelevant to questions about whether group selection is operating. Neighborhood and contextual analysis, for example, are nominalist methods because they do not require that group members be united by physical, social, or biological relations. But neither do these methods exclude such real, unified, groups. Thus for nominalists, as I mean the term, the members of the groups may, but need not, be united by some set of physical, social, or biological relations, and, if they are united, the relevant relations may, but need not, be known to obtain.

At least some biologists, especially those engaged with actually modeling wild populations (see sec. 4) adopt practices that are consistent with nominalism. But many orthodox accounts of group selection are less permissive: it is thought that the collections to which individuals are mapped ought not be mere neighborhoods but rather groups that are real in the sense that members' evolutionary fates must be tied to one another in some interesting and important way. There are roughly three strategies for defining groups so as to induce such partitions: appeal to expert knowledge, appeal some observable social or biological relation among individuals, or appeal to fitness-affecting interactions among individuals. The last is by far the most common strategy, and accordingly we will begin with it.

2. Fitness-Defined Groups. On this method of inducing groups, two individuals are held to be members of a common group (if and) only if they affect one another's fitness. As Sober and Wilson put it, "a group is defined as a set of individuals that influence each other's fitness with respect to a certain trait but not the fitness of those outside the group" (1999, 92), which definition they take, plausibly, to summarize the essential feature of groups as understood by Darwin (1871), Haldane (1932), Wright (1945), Williams and Williams (1957), and Hamilton (1975). Godfrey-Smith (2008) endorses a similar requirement in at least some circumstances and cites in support Uyeno-

yama and Feldman (1980), Wilson (1980), Michod (1982), and Wade (1985). Others endorse some version of the requirement but add further restrictions or allow exceptions of one kind or another. Okasha (2006), for example, distinguishes between ecological and genealogical groups and requires only the former to satisfy the requirement; he also requires that ecological groups be capable of "free living." Sterelny (1996), Maynard Smith (1998), and Nunney (1998) arguably endorse some version of the requirement and also, as Okasha notes, require a richer functional organization among the elements of the group. Thus, for many theorists, symmetric, fitness-affecting causal interactions between (nearly all) members of the group and the absence of such interactions between (nearly all) members of different groups is a necessary feature of group structure: given a mapping G from the domain of individuals to subsets thereof, if the 'groups' identified by the mapping systematically include individuals that do not influence one another's fitness, then the mapping does not identify real groups, and there is no group selection acting on that population with respect to those collections of individuals. I say that theories including this kind of constraint on group structure employ *fitness-defined* groups.

Fitness-defined groups introduce a decidedly intractable discovery problem for those who would model populations using real data. If group selection occurs only over real rather than nominal groups, we can diagnose the occurrence of group selection only given a prior diagnosis of which partitions of the population yield real groups. And to do that one must specify a test for the existence of the group-defining relations, here fitness-affecting interactions. Such tests are complicated by a necessary vagary. Godfrey-Smith (2008) points out, correctly, that often no partition of the population will yield collections of organisms, cells in the partition, each of which strictly satisfies the constraints on fitness-defined groups. But he and others suppose, as will I, that often enough there are partitions on which the resulting collections approximately satisfy the relevant constraints, hence the 'nearly all' in the above framing. I will gloss 'approximately satisfy' thusly: we require a partition of the population such that the network of fitness-affecting interactions is relatively dense within groups and relatively sparse between groups. This is vague in respect of what counts as 'dense' and 'sparse', and there is clearly a disputable boundary: there will be cases in which it is unclear whether there are fitness-defined groups because there are only a few fitness-affecting interactions and such as do occur are only slightly more frequent within than between groups. But also I assume that there are clear limiting cases, that is, cases in which fitness-affecting interactions occur but are so sparse even within groups that on any reasonable interpretation of those who require fitness-defined groups no such groups are present.

Developing tests for the satisfaction of vague boundary conditions is of course problematic. But as it turns out precision here is not to the point. Min-

imally, we want a procedure that, when given a partition, reliably determines whether the collections generated by that partition do or do not satisfy the (now vague) constraints on fitness-defined groups. No procedure of this sort is now possible, exactly because there are no reliable methods for determining the existence of fitness-affecting interactions between individuals.[1] Such interactions involve cross-unit causal dependencies, which turn out to be empirically inaccessible. More precisely, on any reasonable precisification of 'dense' and 'sparse', available methods are not appropriately sensitive to the frequency or distribution of cross-unit fitness-affecting interactions. In fact, there is in the literature just one general strategy for identifying the presence of the relevant fitness-affecting interactions.[2] Before elaborating the challenges it faces, it will be useful to have a clear view of what makes it so difficult to discover cross-unit causal dependencies.

2.1. Inferring Cross-Unit Causation.   The fitness-affecting interactions employed to define real biological groups are a species of cross-unit causation. Cross-unit causation occurs whenever the traits of one individual, or unit, causally influence the traits of another. It is simple enough to define such cross-unit causal dependencies. For instance, on an interventionist conception of causation (Pearl 2000; Spirtes, Glymour, and Scheines 2000; Woodward 2003) it is perfectly reasonable to say than Anya's income causally influences Boris's education and that this is so if and only if there is some intervention on Anya's income that changes the probability density over Boris's education. But discovering the truth about such causal hypotheses is not so simple at all.

The reason for this is that causal dependencies are evidenced by the statistical associations they generate in the data, and (more important) the absence of those causal dependencies is evidenced by the absence of associations in the data. But to find associations in the data, or their absences, variable values must be paired. Sample covariance, for example, is defined as the mean product of paired deviations from the mean. Consider Boris and Anya again. If one wanted to test the theory that income causes education, one might look to see whether there is a sample covariance between the var-

---

1. More exactly: there are tests that, on representative data from a population, determine whether there exists at least one fitness-affecting interaction; there are however no tests that reliably determine for each pair ⟨a,b⟩ in the population whether some trait of a's influences b's fitness.

2. Although see, e.g., Hudgens and Halloran (2012), Aronow and Samii (2013), Ugander et al. (2013), and Gui et al. (2015) for discussions of novel methods for estimating the strength of cross-unit dependencies (so-called interference or indirect effects) in networks, which methods may hold some promise for cross-unit causal inference using experimental data.

iables: where i indexes individuals and n is the number of individuals in the sample, is

$$\frac{1}{n-1} \sum_{1}^{n} \left(\text{Income}(i) - \overline{\text{Income}}\right) \left(\text{Education}(i) - \overline{\text{Education}}\right)$$

nonzero? If so, Income and Education are associated and so potentially causally related; if not, then neither causes the other, and, more, they share no common cause. But calculating the covariance requires that observations of income and education be paired, which is what the index i is doing. Normally, i just indexes subjects, so we pair the observation of Anya's income with the observation of Anya's education and the observation of Boris's income with the observation of Boris's education, and so on. But clearly that will not do to test for cross-unit causation: that Anya's income influences Anya's education and Boris's income influences Boris's education implies nothing about whether Anya's income influences Boris's education. Hence, any test of cross-unit causation requires some other way of pairing observations.

Nor will it do to simply pair Anya's income with Boris's education, for that gives us a sample of one, from which nothing can be inferred with any reliability (e.g., the sample covariance is not even defined on a sample of one). One could simply consider every possible pair, but it is not clear what a nonzero sample statistic from such a data set would mean, and, not unrelatedly, for samples of any size at all, the signal generated by the causal dependency would be swamped. The problem is really twofold. First, observations must be paired or otherwise grouped, unit a with unit b, unit c with d, and so on (a matching problem). Second, within the resulting groups the observations must be ordered, potential cause–potential effect, a's education with b's income, c's education with d's income, and so on (an ordering problem).

There are methods for avoiding the ordering problem, for example, the use of intraclass correlations or demographic variables, but their values provide at best deeply ambiguous evidence for the existence of pairwise causal dependencies.[3] Some of the relevant problems are simply endemic to causal inference. For example, we could match observations of Education and Income by pairing off our population into families; thus, if Anya and Boris are related as spouse to spouse, they would share a common value for an index variable Family. We do not actually need to arbitrarily order spouses to determine whether units within a given family are likely to share similar education values, and we can test for that using intraclass correlations. If the units with a given family are especially likely to have similar Education values, then Family is associated with Education. If so, then either Family in-

3. I thank an anonymous referee for pointing out the need to treat intraclass correlations explicitly.

fluences Education, or they share some common cause, for example, Income or City of Residence. Similarly, if we map units to groups with a function G and the intraclass correlation of unit fitnesses within a group is nonzero, we have reason to think either the units influence one another's fitness or there is some common cause of their fitnesses. The latter possibility is worrisome if we are committed to fitness-defined groups, but the risk of confounder bias is ever present and unavoidable if we wish to learn from observational data. If this were the only worry, we could proceed apace. Unfortunately, it is not. The substantive worry is best illustrated as it arises in the use of demographic variables, which method is by far the most common in discussions of group selection.

2.2. Demographic Variables.   A demographic variable, as I will use the term here, is a variable that records the value of a group variable in the group to which an individual belongs. Thus, for example, if we pair Anya and Boris as a group, and Anya has 12 years of education and Boris 10, they each belong to a group characterized by a mean education of 11 years. We can then define $E(i)$ as the years of education had by unit i, and $D(i)$ as the mean of E among all units belonging to the same group as i (i.e., $\bar{E}(\mathbf{g})$, where $\mathbf{g}$ is the group to which i belongs). Each unit is then characterized by a pair of variable values: one value for E and another for D.[4] It is important to hold clearly in mind that any given demographic variable D is individuated from others by two things: the trait variable it aggregates, here E, and the mapping function that collects units into groups. If either is changed, one has a distinct demographic variable.

The sometimes explicit but often implicit strategy is this: to determine whether the trait value of unit a influences the fitness of unit b, test whether the mean of trait values of units a and b influences the fitness of b. For example, Sober and Wilson write, just after offering the above quoted definition of 'group', that "mathematically the groups are represented by a frequency of a certain trait, and fitnesses are a function of this frequency. Any group that satisfies this criterion qualifies as a group in multilevel selection theory" (1999, 92–93). This kind of test, it turns out, is a disaster for proponents of group selection. To see why, care must be taken here to disambiguate subtly different ideas.

---

4. Formally, if $E(i)$ is measured on units in a set $\mathbf{P}$ of units, and $G(i)$ is a mapping of units to subsets $\mathbf{g}$ of $\mathbf{P}$, and $M(\mathbf{g})$ is some moment of the distribution of E in the subset $\mathbf{g}$, then $D(i) = M(G(i)) = m$ is a demographic variable measured on units in $\mathbf{P}$ but recording the property of being mapped by G to a subset of $\mathbf{P}$ characterized by a distribution over E with moment $M = m$. I here assume that G will partition $\mathbf{P}$, although the extension to neighborhood variables is obvious.

First, note that the same test—a test for the causal influence of a demographic variable on reproductive success—is being used to establish both the reality of groups and the reality of group selection with respect to those groups (at least, if group selection is understood as MLS1 sensu Heisler and Damuth [1987], Damuth and Heisler [1988], and Goodnight, Schwartz, and Stevens [1992]). The double use is described by Okasha (2004) and illustrated by Stevens, Goodnight, and Kalisz (1995), who defend the choice of group boundaries by appeal to the fact that smaller neighborhoods (.5 meters) generate signification regression coefficients (so selection is occurring), but the use of larger neighborhoods does not increase the variance explained by the regression model (so the smaller neighborhoods are correctly sized).[5] But although the two inferences are grounded in the same data and evidenced by the same test statistics (regression coefficients and R^2 values), they are different inferences, and they are differently reliable. As I explain below, it is possible for the demographic variable to cause individual fitness even when most members of the group do not influence each other's reproductive success.

Second, the quoted passage from Sober and Wilson invites two different readings. We might simply hold that what it is for cross-unit fitness-affecting interactions to occur within a set of organisms **g** just is for $D(i)$ to cause fitness $W(i)$. Differently, we might take the existence of a causal dependency between $D(i)$ and $W(i)$ as evidence for, as a signal of, a fitness-affecting interaction between units a and b, when $D(b)$ is calculated with respect to a group containing a and conversely. I will call the first the 'identity assumption' and the second the 'evidential assumption'. Somewhat different problems arise on the alternative understandings.

Given either assumption, the standard test for fitness-influencing interactions imposes pairings by employing demographic variables defined over some set of collections, which are otherwise generated. If the units in each group exhibit a trait T that affects the fitness of others in the group, then a demographic variable aggregating T in the groups will causally influence individual fitness. Thus, to test whether there is group selection, one tests for such a dependence. To do that one maps individuals to collections (which will be counted as groups if the test is passed) and computes the mean of the frequency distribution of the relevant trait variable in each such collection. That value is then recorded as the value of a demographic variable measured on the unit. When groups are real such a value represents the property of belonging to a group characterized by such and such a mean (or variance or whatever) value of the trait in question. Members of the same group will

5. It is unclear whether Stevens et al. should be understood as committed to fitness-defined groups. The use of neighborhood variables suggests they are not; the use of R^2 to justify the choice of neighborhood size suggests they are.

share the same value for such variables, but because the variables are measured on units, covariances and other measures of association can be calculated with paired values induced by the standard index over units in the sample.[6] So the causal dependence between $D(b)$ and $W(b)$ will be signaled by a conditional association between $D(i)$ and $W(i)$, controlling for $T(i)$. And the causal dependence between $D(b)$ and $W(b)$ either entails (on the identity assumption) or is a reliable signal of (on the evidential assumption) a cross-unit causal dependence between a's trait value and b's fitness.

This is the procedure employed in contextual analysis (cf. Mason, Wong, and Entwisle 1983; Heisler and Damuth 1987), and it has much to recommend it. The manner in which pairings are induced is straightforward, and it is statistically fairly easy to accommodate. Alternative statistical methods for hierarchical modeling exist (see, e.g., Raudenbush and Bryk 2002; Gelman et al. 2004). But, importantly, all such methods will depend, directly or indirectly, on fitness-affecting interactions, if they exist, inducing an association between the relevant demographic variable and fitness. However, the signal is decidedly imperfect. Apart from the above noted possibility of unmeasured confounders, it is possible for the demographic variable $D(i)$ to cause $W(i)$ for some but not all units, that is, for $D(i)$ to cause $W(i)$ when some, but not all, groups are real.[7] That would be bad enough, but worse problems loom.

Assume the interventionist conception of causation (e.g., Pearl 2000; Spirtes et al. 2000) or one of the related conceptions (e.g., Rubin 2005) that permits systematic causal inference from observational data. Consider a population **P** of units indexed by i and including units a and b, and let $G(i)$ be a function from **P** to subsets **g** of **P** that partition **P** so that if $G(j) = \mathbf{g}$ then $j \in \mathbf{g}$. Suppose that a trait of individual a, $T(a)$, causally influences reproductive success for b, $W(b)$; that is, there is some intervention on $T(a)$ that changes the probability density over $W(b)$. Units a and b should then be put in the same group. Define $\bar{T}(g)$ as the mean value of T in the set **g**, and define the corresponding demographic variable $D(i) = \bar{T}(G(i))$. Then given that $T(a)$ causes $W(b)$, $D(b)$ causes $W(b)$: the interventions on $T(a)$ that change the probability density over $W(b)$, which must exist because $T(a)$ causes $W(b)$, also change $\bar{T}(\{a, b\})$ and thus $D(b)$. Assuming the Causal Markov and Faithfulness conditions (cf. Spirtes et al. 2000), these causal de-

---

6. When groups are not real, the value of a demographic variable represents the property of being mapped to such a collection rather than belonging to the collection. Godfrey-Smith (2008) thinks such variables represent properties of the organism's environment, and for this reason, among others, neighborhood variables should not be used in models of group selection. Others disagree (cf. Glymour and French 2009).

7. That is, it might be that some, but not all, members of the population belong to fitness-defined groups. Aspects of this untoward possibility are explored by Basl (2011) and McLoone (2015).

pendencies will be signaled in relevant data by appropriate associations.[8] Thus, in any representative sample data $D(i)$ and $W(i)$ will be associated, and our test for cross-unit fitness-affecting interactions will, appropriately, return a positive result.

But now consider the set {a,b,c} for any other unit c in the population, and the mapping G' that yields this set as the image of each of the units a, b, and c. The interventions on $T(a)$ that modify the density over $W(b)$ are no less interventions on $\bar{T}(\{a, b, c\})$ than on $\bar{T}(\{a, b\})$. Hence, the former causes $W(b)$ if the latter does.[9] So given the demographic variable $D'(i) = \bar{T}(G'(i))$, $D'(b)$ causes $W(b)$, and more generally $D'(i)$ causes $W(i)$. Our test will, now inappropriately, return a positive result for the existence of cross-unit fitness-affecting interactions among a, b, and c, even when no trait of c causally influences the fitness of either a or b.

Interestingly, while the causal dependence between $D(i)$ and $W(i)$ will only be evidenced by an association when G partitions the population into multiple groups, it exists even when G maps all individuals in the population to just one group. Let $D(i)$ be defined with respect to mapping G and be such that $D(i)$ and $W(i)$ are associated in virtue of a cross-unit dependency in one or more of the groups generated by G. Let $D'(i)$ be defined with respect to the same trait variable and a mapping G' that maps every member of the population to just one group. It follows from the fact that $D(i)$ causes $W(i)$ that $D'(i)$ causes $W(i)$: the existence of the causal dependency rests only on the availability of interventions, and any intervention on D defined with respect to G is an intervention on D' defined with respect to G'. Hence, if the presence of a causal relation between $D(i)$ and $W(i)$ is taken to be a reliable signal of fitness-affecting interactions, then conditional associations between $D(i)$ and $W(i)$ controlling for $T(i)$, on any mapping G, are sufficient to infer to the existence of one, but only one, fitness-defined group. On any view that defines group membership by fitness-affecting interactions and makes the identity assumption, group selection really does, literally, reduce to frequency-dependent selection.

In practice, biologists do not make the relevant inferences. Those who employ contextual analysis are generally happy enough with any mapping G on which there is a significant association between $D(i)$ and $W(i)$ explain-

8. For example, a nonzero partial regression coefficient for $W(i)$ when regressed on $D(i)$ conditional on $T(i)$.

9. The causal dependency is implied unless by adding c to the group we intervene on or otherwise perfectly compensate for the $T(a) \to W(b)$ dependency. So, e.g., perhaps $T(a)$ influences $W(b)$ when, but only when, c is not a member of the group, so that an intervention on $\bar{T}(\{a, b\})$ changes the density over $W(b)$, but an intervention on $\bar{T}(\{a, b, c\})$ does not. But in that case there really are fitness-affecting interactions among a, b, and c, and they really do belong in the same group.

ing any appreciable fraction of the variance in $W(i)$, and (as in Stevens et al. 1995) associations weaken as group size increases.[10] This suggests that the evidential assumption is the more charitable reading of Sober and Wilson. But in thus avoiding universal groups, practice becomes flatly inconsistent with the requirement that groups be fitness defined. Fitness-defined groups require dense pairwise fitness-affecting interactions within groups, but the existence of causal connections between $D(i)$ and $W(i)$ is really bad evidence for such dense within-group interactions. To see this, simply consider mappings that include several units in each of the groups but include in each group just one pair with a cross-unit dependence between them, and let that interaction be very strong. Then, although by assumption most units in most groups do not influence one another's fitness, $D(i)$ will both cause and, assuming the Causal Markov and Faithfulness conditions, be associated with $W(i)$.

One might attempt alternative methods. For example, one could engage in pairwise testing of every ordered pair $\langle i, j \rangle$ of individuals in the population and then infer the group structure by employing one or another clustering algorithm (see, e.g., White and Reitz [1983] for early efforts, Clauset, Newman, and Moore [2004] or Newman and Girvan [2004] for modularity based methods, Ding et al. [2001] for normalized cut methods, among other alternatives) on the resulting network to construct an approximate partition. The problems here are twofold. First, with n units one has n observations to test $n^2 - n$ hypotheses ($\bar{T}(\{i, j\}) \rightarrow W(j)$ for each ordered pair $\langle i, j \rangle$ of units in the population, $i \neq j$), and so for $n > 2$ we will need to perform more tests than we have data points. And second, as noted above, for each particular pairing $\{i, j\}$ we have exactly one observation (for the pair $\langle i, j \rangle$ the observed pair of $\bar{T}(\{i, j\})$ and $W(j)$). That strategy commits to statistical nonsense twice over.[11]

2.3. How Serious Is the Problem?   Recapitulating the argument so far: to test for the presence of group selection or to model the selection pressures acting on a population in ways that distinguish multilevel selection processes from individual level selection processes we require a mapping of units to groups. A (perhaps the) standard theoretical presupposition is that these groups must be real and that real groups are fitness defined; that is, real groups are characterized by a network of cross-unit fitness-affecting interactions that is dense within groups and sparse between groups. Because it is not possible

10. I thank an anonymous referee for pointing out the need to make this point explicitly.

11. I note in passing that the above problems are equally applicable to any bit of conceptual analysis, of which there are many in philosophy of biology, that analyzes some concept in terms of cross-unit causation. The propriety of any application to data of the resulting concept is epistemically inaccessible.

to reliably test for pairwise causal dependencies between each (or any) pair of units, the presence of fitness-affecting interactions is understood to be signaled by the casual influence of demographic variables on unit fitnesses. That identification can be definitional or evidential.

On the definitional assumption, the causal dependence between $T(a)$ and $W(b)$ is said to be identical to the dependence between $D(b)$ and $W(b)$, where D is defined relative to a function G that maps a and b to the same group. Hence, tests for the latter are tests for the former. This turns out to yield the disastrous consequence that if there is any unit u such that $T(u)$ causes $W(b)$, then for any unit i, $T(i)$ causes $W(b)$. In consequence, there is either one group in the population or none. On the (more plausible) evidential assumption that $T(a)$ causes $W(b)$ is not identical to $D(b)$ causing $W(b)$, the latter causal dependency is taken to be a reliable signal of the former, and hence tests for the latter are tests for the former. But, absent a definitional identification of the two dependencies, the causal dependency between $D(b)$ and $W(b)$ is in fact not a reliable signal of the causal dependency between $T(a)$ and $D(b)$ in that it will yield false positive verdicts. In particular, when there are strong but sparse within-group fitness-affecting interactions and no (or weak and rare) between-group interactions, available methods will wrongly diagnose the reality of groups. Thus, in precisely the limiting cases described in section 2 in which we are most in need of a reliable method, none are to be had.

One might seek to defend the evidential reading by noting that data are rarely perfect, and when confronted with large groups with sparse within-group networks of interactions, the associations between $D(i)$ and $W(i)$ will often be undetectable, while small groups will tend to have dense networks of within-group interactions. And it is true that the performance of contextual analysis as a method for identifying fitness-defined groups depends essentially on the causal system governing fitness, and in particular on the structure of the causal dependencies and their relative strengths and signs. For some systems, the method will work well, although for others it will not. But it is at best unduly optimistic to assume that most of the time the systems of interest are such as to permit contextual analysis to work well. There is something decidedly untoward about diagnosing a condition (dense within-group fitness-affecting interactions) by adopting a test (associations between $D(i)$ and $W(i)$) for an unreliable indicator (causal dependencies between $D(i)$ and $W(i)$) of the condition, when one knows the test is sensitive to unrepresentative data, and then justifying the choice on the hope that the data will be unrepresentative in just the way required for the unreliability in the test to mitigate the unreliability of the indicator.[12]

---

12. I thank an anonymous referee for pointing out the need to make this argument explicitly.

Differently, one might defend the evidential understanding on the grounds that the imputed unreliability is no different from that affecting any diagnosis of selection. Suppose one tests for selection on trait variable $T(i)$ by estimating the relevant selection gradient, which estimate turns out not to be significantly different from zero. One is then entitled to infer that there is no selection on $T(i)$ but not that there is no selection on the population. Just so, it might be thought, if one tests for group selection by estimating a selection gradient on $D(i)$, defined with respect to mapping $G(i)$, which estimate turns out not to be significantly different from zero, then one is entitled to infer that there is no group selection on $D(i)$ but not that there is no group selection at all.[13]

This objection has some initial plausibility. Certainly, we do not test for selection per se by estimating selection gradients; rather, one calculates population genetic parameters whose values, singly or in comparison, indicate the operation of selection: is the population in Hardy-Weinberg equilibrium, what is Tajima's D, what is the ratio of effective population size to neutral mutation rate, and so on. But, when the regression of $W(i)$ on $D(i)$ yields a coefficient not different from zero, the inference we make is sometimes that group selection is not acting on the population: the assumption is that group selection does not act, and it stands as the 'null' hypothesis unless conclusively refuted.

Moreover, there may be no selection on $D(i)$ either because $G(i)$ yields the wrong grouping or because aggregations of $T(i)$ are causally irrelevant to $W(i)$. If we infer the absence of group selection on $D(i)$ from a failure of $G(i)$ to yield fitness-defined groups, it is important not to also infer that no aggregation of $T(i)$ is relevant to fitness. Putting the point somewhat differently, for those modeling selection using multilevel models for predictive or explanatory ends, the question is not so much whether group selection is acting but whether the measured advantage for some values of $T(i)$ over others is in part due to, or is in part counteracted by, the mean of $T(i)$ in some collection, that is, whether a good model of the population's behavior in respect of $T(i)$ and $W(i)$ over evolutionary time will need to be hierarchical, and if so how best to specify such a model. And to answer these questions it will not do to infer from the absence of fitness-defined groups to the conclusion that hierarchical models are unnecessary. Such inferences are not uncommon among critics of group selection. To take just one recent example, Grinsted, Bilde, and Gilbert (2015) challenge a study by Pruitt and Goodnight (2014) claiming to show group selection in a social spider. Grinsted et al. argue, inter alia, that there are no groups because there is no evidence of within-group fitness-affecting interactions, writing: "The chosen species, *Anelosimus studiosus*, is solitary, rarely forms groups, and shows no evidence

---

13. A version of this objection was advanced by an anonymous referee.

of reproductive restraint or skew within groups" (2015, E1). They then immediately infer that no hierarchical modeling is required, writing: "Both predictions of Pruitt and Goodnight could follow equally well from individual-level selection as from group selection. . . . Merely demonstrating differential survival of groups does not allow the authors to distinguish successful groups from groups of successful organisms" (E1). The latter claim is true and it is relevant if group selection, and thus the need for multilevel models, depends on the existence of fitness-defined groups. But the point is also irrelevant to the question of whether hierarchical (e.g., group selection) models are required for optimal prediction: a model in which both $T(i)$ and $D(i)$ are used to predict fitness may perform better than a model in which $T(i)$ alone is used, and it may do so exactly because $D(i)$ causes fitness; this is possible even when $D(i)$ is defined with respect to a mapping $G(i)$ that does not yield fitness-defined groups, as, for example, when fitness-affecting interactions are strong but obtain between only a sufficiently small minority of group members.

Here then is the most important reason of all to be clear about the limits of inferences from the fact that demographic variables cause fitness to the reality of fitness-defined groups. If those who wish to seriously entertain the possibility of group selection agree that group selection requires fitness-defined groups but can in any study provide as evidence of such groups only a measured association between $D(i)$ and $W(i)$ or the equivalent, they open themselves to legitimate objections. Such associations are sometimes strong evidence that $D(i)$ causes $W(i)$. But they are simply not good evidence that members of the groups generated by the mapping $G(i)$ with respect to which $D(i)$ is defined are characterized by a high density of pairwise fitness-affecting interactions. Work that commits to fitness-defined groups but employs standard techniques to test for their existence rests on bad method.

3. Groups Otherwise Defined. Groups need not be fitness defined. One might instead define groups by some other real physical, biological, or social relation and so without explicit reference to fitness at all. In particular, if one has some prior commitment to some particular mapping, the worries about proper specification may reduce to worries about whether demographic variables defined on that mapping are causally relevant to the effect of interest. Suppose, for example, that Anya is Boris's mother, and one is in particular interested in whether one's mother's income or one's father's income has a greater influence on education. Then one might simply define the variable Mother's Income, measured on individuals. If one can for each individual identify a mother and her income, one can proceed again by using the index i over units to pair observations in order to calculate relevant sample statistics.

This method has several advantages. Since groups are no longer fitness defined, we need not employ tests for an association between demographic variables and fitness as (unreliable) tests for fitness-affecting interactions

but rather as (at least sometimes reliable) tests for existence of a causal connection between the demographic variable and fitness (i.e., as tests for group selection). Further, the range of methods by which to identify groups is considerably expanded. For example, some causal interactions, generally social but sometimes biological, are directly observable: grooming, mating, feeding, and so on are pairwise interactions that can be seen. Even though whether the interaction in turn effects fitness cannot simply be observed, the fact of the interaction can be. Collections of such interactions constitute a social or biological network with various structural features that can be used to partition a population into groups. Differently, one might, as with mothers or more generally with families, identify groups simply by adverting to reasonably well understood features of social life. One could with equal ease appeal to more narrowly held expert knowledge of particular collections—for example, one could identify baboon troops or lion prides by appeal to the expert knowledge of field biologists observing the baboons or lions; certainly the identification of nests, colonies, clutches, and like quite often proceeds by appeal to such expert knowledge. Differently again, one could represent observations of mating, feeding, grooming, or the like with a graphical model and then deploy some clustering algorithm on that model to produce a partition of the population. Herbers and Banschbach (1999), for example, employ several of these strategies when they individuate 'nests' and 'colonies' by appeal to pairwise behavioral tests, spatial location, and genetic data.

On any of these ways of partitioning a population into groups, one can then sensibly seek to test whether demographic variables defined on that partition causally influence reproductive success, although it should be said that the statistics here are seriously nontrivial and clustering is more a matter of art than science. We should recognize that even setting aside statistical concerns this strategy has certain disadvantages. First, the resulting groups are not fitness defined. Groups are rather defined by whatever physical, social, or biological relation is employed individuating them, which relations will themselves often be unknown when one appeals to expert knowledge. Second, while there is nothing essentially wrong with an appeal either to expert knowledge or widely shared common knowledge, the best defense of such everyday or expert identification of groups invites a nominalism about groups.

4. Nominalism Again. Suppose we defend a particular partition of baboons into troops by appeal to the fact that the experts, the field biologists who spend time actually observing the baboons, recognize just those groups. Even if we did not bother to build a graphical model of the grooming, mating, display, and so on, behaviors or bother to use a clustering algorithm to induce a partition of the population into troops, we could reasonably rest con-

tent with the resulting partition of the population. This is because we can be reasonably sure that any good clustering algorithm applied to such a model would generate the same groups as the field-workers recognize. And that is so because if a clustering algorithm did not routinely return just the clusters recognized by the experts, we would reject the clustering algorithm as bad news: whatever it is that such algorithms are identifying, it is not the groups of interest to us, however objective or real they might be. But if this is the proper defense of an appeal to expert knowledge, and that appeal is legitimate, nominalism about groups seems entirely appropriate. The right groups to consider are whatever groups happen to interest us at the moment.

This of course implies that diagnoses of the presence or absence of group selection are necessarily relative to an arbitrary, although not unmotivated, partition of the population into groups. But at least this aspect of nominalism seems not only harmless but correct: if there are in fact cross-unit fitness-affecting causal dependencies, it will be true that on some partitions of the population demographic variables influence fitness; this will be true, for example, for those partitions that yield collections satisfying the constraints of fitness-defined groups. But there often will be other partitions on which demographic variables do not cause fitnesses; this will be true when the collections generated by the partition never include individuals that affect one another's fitness. Insofar as MLS1 versions of group selection are, at least in part, a matter of demographic variables causally influencing individual reproductive success, judgments about whether group selection is actually occurring really ought be relative to the partition of the population employed. That the implicit relativization is made explicit by nominalism then looks to be a feature rather than a bug.

Nominalism, as I am here using the term, is anathema to many, perhaps most, of the standard classical discussions of group selection and much of the theoretically motivated commentary thereon. In that tradition, groups must be real, and reality is a matter of fitness-affecting interactions. Individuals come as units of a group, it is thought, when, and only when, to some appreciable extent the evolutionary fate of each individual is bound up with that of the others. But experimental and observational work, especially that in the traditions following Heisler and Damuth on the one hand or Wade on the other, is often less demanding. Some discussions seem to presuppose the importance of fitness-affecting interactions. For example, Pruitt and Goodnight, writing in defense of an earlier paper (Pruitt and Goodnight 2014) in which they claim to have demonstrated group selection in a social spider, seem to think it is important that individuals in a colony succeed or fail together, that is, that the behavioral or social interactions used to identify groups also constitute fitness-affecting interactions. They write, "Our case study is clear because both the target and agent of selection are above the level of the

individual: the target of selection (group composition) is a trait that an individual cannot have, and the agent of selection (extinction) is the textbook example of strong group selection. We showed that *A. studiosus* colonies live or die as a unit" (Pruitt and Goodnight 2015, E4).

But many other studies seem to employ partitions in which groups are defined by appeal to expert knowledge. Moorad (2013), for example, employs a variable measuring the number of mates for an individual's father, which father is in turn identified from birth records rather than genetic data; thus, an individual's male parent is identified by appeal to records of common-knowledge identifications. Tsuji (1995), for another example, identifies ant colonies (*Pristomyrmex pungens*) with single nest sites, claiming that colonies were monodomous (i.e., had but one nest site), but offers no data in support of that claim (i.e., colonies have been individuated on the basis of Tsuji's expert knowledge). Similarly, Breden and Wade (1989) consider egg clutches, which clutches are distinguished one from another not by data on which an analysis is performed but by expert knowledge. This use of expert knowledge need not be epistemically fraught, but neither are the resulting groups constructed on the basis of fitness-affecting interactions.

Yet other studies are based on physical or biological relations other than fitness (e.g., Herbers and Banschbach 1999; Laiolo and Obeso 2012). Laiolo and Obeso partition a metapopulation of Dupont's lark into local populations following Vögeli et al. (2010), who divide the study location into patches on the basis of bird movement, census data, and habitat. Still other studies individuate groups in apparently arbitrary ways. Aspi and coauthors, studying selection on patches of Tatar catchfly, identify a patch with "a group of individuals within a maximum (arbitrary) distance between individuals of five meters" (2003, 510). Eldakar et al. (2010), in a study of water striders, identify pools and pool regions on an ephemeral stream bed, the latter being a major pool with its "immediately connected" minor pools. No criteria are given for distinguishing 'immediately connected' pools; whatever criteria were used were apparently unmotivated to details of water strider behavior or biology. Weinig and coauthors (2007) designed experimental 'patches' of *A. thaliana* by planting seeds in pots; within each pot seeds were planted on a 3×3 grid, with 1 centimeter between grid locations. Pots are treated as patches, but no test for within-pot interactions of any kind are offered in justification of the choice of 1 centimeter distances between seeds. Again, this is not problematic, unless we require that groups be fitness defined.

Discussions in empirical papers often underdetermine the authors' considered views about the nature of the groups required for group selection. On the one hand, all of these studies, excepting Pruitt and Goodnight (2015), are consistent with nominalism; in none of them is group identification explicitly made definitionally dependent on fitness-affecting interactions. Per-

haps, then, the authors are perfectly satisfied with groups defined by real physical, social, or biological relations, even if those relations are not pairwise fitness-affecting interactions. Or perhaps the authors simply care about the groups they use, whether or not they have some underlying biological reality. On the other hand, it might be that the authors are laboring under the mistaken assumption that a causal connection between demographic variables and fitness suffices to establish that the groups on which the variable is measured are unified by dense networks of fitness-affecting interactions. Indeed, even pieces like Pruitt and Goodnight (2015) are open to alternative interpretations. Perhaps Pruitt and Goodnight are committed to fitness-affecting interactions. But perhaps they are not and are instead simply responding to a critic who is. But at least this much can be said. The methods employed in these studies do not in fact demonstrate that the studied groups satisfy the conditions on fitness-defined groups. To the extent that the fitness-defined groups are a necessary condition of group selection, the studies simply do not demonstrate group selection. Hence, to the extent that the observational, experimental, and inferential practices therein are acceptable, fitness-defined groups are not necessary conditions of group selection, and, what is more, nominalism about groups looks to be perfectly acceptable.

I make one last observation. Insofar as one is willing to be nominalist about groups (i.e., to hold that the groups identified in a MLS1 model of a population need be no more real than that the individuals assigned to groups really do exist), much of the motivation for preferring group to neighborhood variables vanishes, and there is rather less reason to object to the use of demographic variables defined over neighborhoods rather than groups. Of course, for some that will be sufficient reason to reject nominalism. But those who do owe a fuller, more careful story about just how real groups are to be identified and why reality, so understood, is essential.

5. Implications. The situation then appears to be this. If we reject nominalism, we require real groups. The standard account of such groups is that any two individuals in a population belong to the same group if and only if there are fitness-affecting causal dependencies between them; that is, for some trait T, either $T(a)$ causes $W(b)$ or $T(b)$ causes $W(a)$. Such dependencies, it turns out, can be tested for, given current methods, only by testing for a causal dependence between individual fitness and a demographic variable $D(i) = \bar{T}(G(i))$, where G maps individuals to groups. That method is seriously unreliable. Causal dependencies between demographic variables and fitness can, indeed will, hold even when most members of most groups do not influence one another's fitness, and the associations that are evidence for the causal dependence between demographic variables and fitness can hold even when the mapping $G(i)$ generates groups that are not characterized by a high density of pairwise fitness-affecting interactions.

There are alternatives that preserve something of the flavor of the initial proposal. We might define groups by means of causal dependencies between variables other than reproductive success and then ask whether these real groups are such that demographic variables defined over those groups causally influence the reproductive success of the members. When the relevant interactions are themselves directly observable, methods for using the network of such relations to partition the population are available, and, depending on the way in which group structure is induced, raise only manageable statistical problems. But in adopting these methods we must relinquish the two most important features of groups used to argue against nominalism about groups. First, the idea that groups should include only individuals that affect one another's fitness must go, for even when it is true that demographic variables influence the reproductive success of individuals, it will not in general be true that the trait value of any one group member influences the reproductive success of any other; indeed it may well be the case that for most groups most members do not influence one another's fitness. Second, although in some sense the group structure is understood to be generated by some underlying network of biological or social relations, we will often not infer a group structure from some representation of the underlying social or biological relations. Indeed, we will often neither represent the relations nor infer them from data but instead directly induce the group structure by appeal to either expert (e.g., baboon troops) or common (e.g., families) knowledge. But the best justification for such appeals invites nominalism insofar as it relativizes judgments about group selection to partitions of the population into groups and justifies the use of one rather than another partition by appeal to our interest in the resulting groups.

Finally, there remains nominalism, which offers a ready justification both for appeal to expert knowledge in inducing group structure and for the use of social or biological causal relations apart from those in which fitness enters as an effect variable. Groups are where you find them, and if your interest is in just these groups, either because those are the groups you care about after having watched Anya and Boris or the baboons or the oak trees, for many years, or because the observed patterns of social interaction generate just those clusters, or for whatever other reason, then the group selection hypotheses of interest will be hypotheses about the influence of demographic or aggregate variables measured on just those groups. As a nominalist who prefers neighborhoods to groups, I recommend this strategy, although the reader's tastes may differ. But whatever one's tastes, if we are to take inference seriously then neither definitions of nor tests for group selection may require that groups be defined by fitness-affecting causal dependencies among group members: to do so places the reality of groups, and hence the reality of group selection, beyond the reach of even our best methods and so also beyond the epistemic pale.

REFERENCES

Aronow, Peter M., and Cyrus Samii. 2013. "Estimating Average Causal Effects under Interference between Units." arXiv preprint. https://arxiv.org/pdf/1305.6156.pdf.

Aspi, Jouni, Anne Jäkäläniemi, Juha Tuomi, and Pirkko Siikamäki. 2003. "Multilevel Phenotypic Selection on Morphological Characters in a Metapopulation of *Silene tatarica*." *Evolution* 57:509–17.

Basl, John. 2011. "The Levels of Selection and the Functional Organization of Biotic Communities." PhD diss., University of Wisconsin.

Breden, Felix, and Michael J. Wade. 1989. "Selection within and between Kin Groups of the Imported Willow Leaf Beetle." *American Naturalist* 134:35–50.

Clauset, Aaron, Mark E. J. Newman, and Christopher Moore. 2004. "Finding Community Structure in Very Large Networks." *Physical Review* E 70:066111.

Damuth, John, and I. Lorraine Heisler. 1988. "Alternative Formulations of Multilevel Selection." *Biology and Philosophy* 3:407–30.

Darwin, Charles Robert. 1871. *The Descent of Man and Selection in Relation to Sex*. London: Murray.

Ding, Chris H. Q., Xiaofeng He, Hongyuan Zha, Ming Gu, and Horst D. Simon. 2001. "A Min-Max Cut Algorithm for Graph Partitioning and Data Clustering." In *IEEE International Conference on Data Mining 2001: Proceedings*, ed. Nick Cercone, Tsau Y Lin, and Xindong Wu, 107–14. Los Alamitos, CA: IEEE Computer Society.

Eldakar, Omar Tonsi, Michael J. Dlugos, Galen P. Holt, David Sloan Wilson, and John W. Pepper. 2010. "Population Structure Influences Sexual Conflict in Wild Populations of Water Striders." *Behaviour* 147:1615–31.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. 2nd ed. Boca Raton, FL: Chapman & Hall.

Glymour, Bruce, and Christopher French. 2009. "Causation, Equivalence and Group Selection." Unpublished manuscript, Kansas State University.

Godfrey-Smith, Peter. 2008. "Varieties of Population Structure and the Levels of Selection." *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* 140:743–61.

Grinsted, Lena, Trine Bilde, and James D. J. Gilbert. 2015. "Questioning Evidence of Group Selection in Spiders." *Nature* 524: E1–E3.

Gui, Huan, Ya Xu, Anmol Bhasin, and Jiawei Han. 2015. "Network a/b Testing: From Sampling to Estimation." In *Proceedings of the 24th International Conference on World Wide Web*, ed. Aldo Gangemi, 399–409. New York: ACM.

Haldane, J. B. S. 1932. *The Causes of Evolution*. London: Longmans.

Hamilton, William D. 1975. "Innate Social Aptitudes of Man: An Approach from Evolutionary Genetics." In *Biosocial Anthropology*, 133–55. London: Malaby.

Heisler, I. Lorraine, and John Damuth. 1987. "A Method for Analyzing Selection in Hierarchically Structured Populations." *American Naturalist* 130:582–602.

Herbers, Joan M., and Valerie S. Banschbach. 1999. "Plasticity of Social Organization in a Forest Ant Species." *Behavioral Ecology and Sociobiology* 45:451–65.

Hudgens, Michael G., and M. Elizabeth Halloran. 2012. "Toward Causal Inference with Interference." *Journal of the American Statistical Association* 103:832–42.

Laiolo, Paola, and José Ramón Obeso. 2012. "Multilevel Selection and Neighbourhood Effects from Individual to Metapopulation in a Wild Passerine." *PloS One* 7: e38526.

Mason, William M., George Y. Wong, and Barbara Entwisle. 1983. "Contextual Analysis through the Multilevel Linear Model." *Sociological Methodology* 14:72–103.

Maynard Smith, John. 1998. "The Origin of Altruism." Review of *Unto Others*, by Elliott Sober and David S. Wilson. *Nature* 393:639–40.

McLoone, Brian. 2015. "Some Criticism of the Contextual Approach, and a Few Proposals." *Biological Theory* 10:116–24.

Michod, Richard E. 1982. "The Theory of Kin Selection." *Annual Review of Ecology and Systematics* 13:23–55.

Moorad, Jacob A. 2013. "Multi-Level Sexual Selection: Individual and Family-Level Selection for Mating Success in a Historical Human Population." *Evolution* 67:1635–48.

Newman, Mark, and Michelle Girvan. 2004. "Finding and Evaluating Community Structure in Networks." *Physical Review* E 69:026113.

Nunney, Leonard. 1998. "Are We Selfish, Are We Nice, or Are We Nice because We Are Selfish?" Review of *Unto Others*, by Elliott Sober and David S. Wilson. *Science* 281:1619–21.

Okasha, Samir. 2004. "Multi-Level Selection, Covariance and Contextual Analysis." *British Journal for the Philosophy of Science* 55:481–504.

———. 2006. *Evolution and the Levels of Selection*. Oxford: Oxford University Press.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Pruitt, Jonathan N., and Charles J. Goodnight. 2014. "Site-Specific Group Selection Drives Locally Adapted Group Compositions." *Nature* 514:359–62.

———. 2015. "Pruitt and Goodnight Reply." *Nature* 524: E4–E5.

Raudenbush, Stephen W., and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Vol. 1. Thousand Oaks, CA: Sage.

Rubin, Donald B. 2005. "Causal Inference Using Potential Outcomes." *Journal of the American Statistical Association* 100:322–31.

Sober, Elliott, and David S. Wilson. 1999. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge, MA: Harvard University Press.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction and Search*. 2nd ed. Cambridge, MA: MIT Press.

Sterelny, Kim. 1996. "The Return of the Group." *Philosophy of Science* 63:562–84.

Stevens, Lori, Charles J. Goodnight, and Susan Kalisz. 1995. "Multilevel Selection in Natural Populations of *Impatiens capensis*." *American Naturalist* 145:513–26.

Tsuji, Kazuki. 1995. "Reproductive Conflicts and Levels of Selection in the Ant *Pristomyrmex pungens*: Contextual Analysis and Partitioning of Covariance." *American Naturalist* 146: 586–607.

Ugander, Johan, Brian Karrer, Lars Backstrom, and Jon Kleinberg. 2013. "Graph Cluster Randomization: Network Exposure to Multiple Universes." In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 329–37. New York: ACM.

Uyenoyama, Marcy, and Marcus W. Feldman. 1980. "Theories of Kin and Group Selection: A Population Genetics Perspective." *Theoretical Population Biology* 17:380–414.

Vögeli, Matthias, David Serrano, Fernando Pacios, and José L. Tella. 2010. "The Relative Importance of Patch Habitat Quality and Landscape Attributes on a Declining Steppe-Bird Metapopulation." *Biological Conservation* 143:1057–67.

Wade, Michael. 1985. "Soft Selection, Hard Selection, Kin Selection and Group Selection." *American Naturalist* 125:61–73.

Weinig, Cynthia, Jill Johnston, Charles Willis, and Julin Maloof. 2007. "Antagonistic Multilevel Selection on Size and Architecture in Variable Density Settings." *Evolution* 61:58–67.

White, Douglas R., and Karl P. Reitz. 1983. "Graph and Semigroup Homomorphisms on Networks of Relations." *Social Networks* 5:193–234.

Williams, George C., and Doris C. Williams. 1957. "Natural Selection of Individually Harmful Social Adaptations among Sibs with Special Reference to Social Insects." *Evolution* 11:32–39.

Wilson, David Sloan. 1980. *The Natural Selection of Populations and Communities*. Menlo Park, CA: Benjamin/Cummins.

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

Wright, Sewall. 1945. "Tempo and Mode in Evolution: A Critical Review." *Ecology* 26:415–19.