Biol Philos
DOI 10.1007/s10539-013-9378-y

Pluralism in evolutionary controversies: styles and averaging strategies in hierarchical selection theories

**Rasmus Grønfeldt Winther · Michael J. Wade ·
Christopher C. Dimond**

Received: 17 October 2012 / Accepted: 12 April 2013

Abstract Two controversies exist regarding the appropriate characterization of hierarchical and adaptive evolution in natural populations. In biology, there is the Wright–Fisher controversy over the relative roles of random genetic drift, natural selection, population structure, and interdemic selection in adaptive evolution begun by Sewall Wright and Ronald Aylmer Fisher. There is also the Units of Selection debate, spanning both the biological and the philosophical literature and including the impassioned group-selection debate. Why do these two discourses exist separately, and interact relatively little? We postulate that the reason for this schism can be found in the differing focus of each controversy, a deep difference itself determined by distinct general styles of scientific research guiding each discourse. That is, the Wright–Fisher debate focuses on *adaptive process*, and tends to be instructed by the *mathematical modeling style*, while the focus of the Units of Selection controversy is *adaptive product,* and is typically guided by the *function style*.

Electronic supplementary material The online version of this article (doi: 10.1007/s10539-013-9378-y) contains supplementary material, which is available to authorized users.

R. G. Winther (✉)
Department of Philosophy, University of California, Santa Cruz, 1156 High Street, Cowell
Academic Services room 224, Santa Cruz, CA 95064, USA
e-mail: rgw@ucsc.edu

M. J. Wade
Department of Biology, Indiana University, 1001 East 3rd St, Bloomington, IN 47405, USA
e-mail: mjwade@indiana.edu

C. C. Dimond
Center for Biology and Society, Arizona State University, SOLS Graduate Programs,
P.O. Box 874601, Tempe, AZ 85287-4601, USA
e-mail: christopher.dimond@asu.edu

The differences between the two discourses can be usefully tracked by examining their interpretations of two contested strategies for theorizing hierarchical selection: *horizontal* and *vertical* averaging.

Keywords Hierarchical selection · Sewall Wright · R.A. Fisher · Units of selection · Levels of selection · Styles of scientific research · Models · Adaptation · Averaging strategies · Averaging fallacy · Group selection

Introduction

The basics

Two controversies exist regarding the appropriate characterization of hierarchical and adaptive evolution in natural populations. In evolutionary biology, there is the Wright–Fisher debate over the relative roles of random genetic drift, natural selection, population structure, and interdemic selection in adaptive evolution. This debate was begun by Sewall Wright and Ronald Aylmer Fisher, and carries over to the present day (e.g. Coyne et al. 1997, 2000; Wade and Goodnight 1998; Goodnight and Wade 2000). There is also a controversy about the units of selection that spans both the biological and the philosophical literature and includes the well-known group-selection debate. Important concerns of the Units of Selection debate include how often selection acts on groups or any other level(s) in the biological hierarchy above the individual, and whether higher-level selection has the appropriate causal structure to produce group-level adaptations.

There is scant cross-referencing between these two controversies. We document and try to explain this schism. The differing focus of each discourse, determined by distinct general styles of scientific reasoning (e.g., Hacking 2002, 2009; Elwick 2007; Winther 2006c, 2012, 2013), goes a long way towards explaining the non-overlap. That is, the Wright–Fisher debate focuses on *adaptive process*, and is usually instructed by the *mathematical modeling style*, while the Units of Selection debate highlights *adaptive product,* and tends to be guided by the *function style*. Analyzing these overarching styles might also assist in classifying and negotiating the aims and methods of other controversies within evolutionary biology (e.g., between sociobiology and population genetics reviewed in Schwartz 2002; evolutionary constraints as seen by game theory relative to evolutionary quantitative genetics reviewed in Hammerstein 1996; or arguments between mathematical evolutionary theorists and mechanistic evolutionary developmental biologists, see Winther 2013). In particular, we identify an important place where the differences in goals and techniques between the two debates are expressed: discussions about the validity of averaging strategies. We introduce a novel distinction between two kinds of averaging strategies: *horizontal* (across the fitnesses of background genes) and *vertical* (across the fitnesses of units at different levels).

Our paper has descriptive, explanatory, and normative aims. Descriptively, we characterize similarities and differences between the Wright–Fisher and Units of

Selection debates not previously identified, such as their respective ways of using averaging strategies. Explanatorily, we suggest reasons for systematic differences, such as adoption of distinct styles of modeling and of explanation. In other words, we attempt to provide a philosophical explanatory gloss of the scant overlap between these two controversies. Moreover, precisely by making style pluralism evident, we contribute to the communal effort of understanding controversies—and potentials for negotiation and integration—in evolutionary theory and beyond. Even if the reader finds our explanation unconvincing, our descriptive goal can succeed independently of our explanatory aim. Normatively, we believe the two literatures and their respective styles should be integrated for some issues (e.g., when can we legitimately locate selective processes at a higher level of selection?), while for other, specialized matters they should remain independent.

What is at stake? Style pluralism in evolutionary theory

The last decade has seen the reinvigoration of a century-long debate between RA Fisher's and Sewall Wright's perspectives on the primary forces acting during the evolutionary process (Coyne et al. 1997, 2000; Wade and Goodnight 1998; Goodnight and Wade 2000). These two views agree on the general mathematical description (i.e., dynamical laws and parameter space; see Lewontin 1974) of the evolutionary process but disagree about the size of the relevant parameter space, and about the importance of population subdivision, interdemic selection, and gene interaction.

A separate debate has focused primarily on the units of selection in relation to adaptation. Here a key goal is to determine whether there are characters or traits best viewed as group-level adaptations. If so, this confirms the existence of group selection, just like functional analysis of phenotypic traits of individuals provides support for individual selection in producing adaptations of individuals (see Gould and Vrba 1982; Hempel 1994).

What are the relationships between these historically distinct, yet related, debates? Are the Wright–Fisher and Units of Selection debates the *same* controversy (the latter perhaps being a philosophical reconstruction of the former) or are they substantively different debates? If the former, then why has the overlap and mutual cross-referencing between them been so miniscule, as we document below? If the latter, then where do the differences reside? What could the philosophy of biology concerned with the Units of Selection debate learn from the Wright–Fisher debate, and vice versa for biology? What could evolutionary theory in general learn from both controversies?

We argue that the two debates are similar in some respects, but are guided by fundamentally different aims, themselves partly determined by styles of inquiry. A key goal of the Wright–Fisher debate is to articulate realistic and complex mathematical-formal genetic models for evolutionary change under a variety of simultaneously acting forces, including natural selection, random genetic drift, migration, and mutation. Natural selection is often a privileged force in this debate because the motivation for the mathematical enterprise arises from an interest in explaining the causes of phenotypic adaptations of individuals. In contrast, an essential motivation underlying the Units of Selection controversy is clarifying the concept of “adaptation for… ”. Natural selection is simply assumed to be the primary force changing gene frequencies; inquiry revolves around the nature of adaptation, the outcome(s) of the adaptive process.

The aims of the two debates are not necessarily in conflict. The functional notion of “adaptation for…” requires an appropriate application of the domain within the Wright–Fisher debate wherein natural selection is the predominant evolutionary force. Moreover, the Units of Selection debate has not incorporated complex genetics (e.g., gene interactions or epistasis) because, for the most part, the underlying genetics of adaptation are irrelevant to discerning the adaptive function of phenotypic traits.

Indeed, an understanding of evolutionary theory must include both complex genetics (with roots in August Weismann and T.H. Morgan) and the “adaptation for…” functional locution (with roots in Darwin) and it must integrate these concepts across various levels of biological organization. Negotiation and possible integration between the two controversies would be helpful. We also argue that these debates use two kinds of averaging strategies: *horizontal averaging* averages over gene–gene and gene–environment interactions to determine genic effects; and *vertical averaging* averages across levels in the biological hierarchy. Finally, we also note that evolutionary biology requires two distinct *styles* of explanation, modeling, and inference: *mathematical modeling*, which guides the Wright–Fisher debate; and *function*, which instructs the Units of Selection discourse. The function style itself requires two further styles: *history* and *mechanism.* History and mechanism are necessary but not sufficient for the function style, because concerns with “adaptation for…” must also be added.

Evolutionary biology is a pluralistic explanatory and predictive enterprise (e.g., Mitchell 2003; Kellert, Longino and Waters 2006). Understanding its style pluralism is necessary for comprehending the richness of its practices and aims. Styles of scientific research are very general ways of doing science, of “finding things out” (Hacking 2009). The historian A.C. Crombie introduced the concept:

> The scientific movement brought together in its common restriction to answerable questions a variety of styles of scientific argument, of scientific methods of inquiry, demonstration and explanation, diversified by their subject-matters, by general conceptions of nature, by presuppositions about scientific validity and cogency, and by scientific experience of the interaction of programmes with realizations (Crombie 1994, vol 1, 83).

Hacking, who has developed the concept over the last three decades, notes:

> Every style of reasoning introduces a great many novelties including new types of: objects; evidence; sentences, new ways of being a candidate for truth or falsehood; laws, or at any rate modalities; possibilities. One will also notice, on occasion, new types of classification and new types of explanations (Hacking 2002, 189).

Following these descriptions, styles provide overarching theoretical and experimental ways of doing science, and of viewing objects and processes in nature. The *standard view* of styles identifies six types: (1) deductive (postulation or axiomatic), (2) experimental, (3) analytical-hypothetical (hypothetical modeling), (4) taxonomic, (5) probabilistic, and (6) evolutionary (historical derivation or genealogy) (Crombie 1994; Hacking 2002, 2009; Kwa 2011; cf. Pickstone 2001).

The styles here used are the following (see Winther 2012, 2013 for further details):

1. *mathematical modeling* style: the analytical-hypothetical "Galilean style" that Edmund Husserl, Noam Chomsky, and Steven Weinberg wrote about, together with probability and statistics. It can be seen as the first and fifth, and probably the third, styles combined around the notion of a mathematical model.
2. *function* style: This style has roots in teleology and notions of design (e.g., Kitcher 1993; Allen et al. 1998; Ariew et al. 2002; Buller 1999). What is a trait or part of a system "good for"? In answering this question, two further styles must be employed:
    a. *mechanism* style: a style essential to biology, thanks to René Descartes, Claude Bernard, and others. It is associated with ubiquitous, and potentially problematic, forms of reductionism. This style is also a particular sort of modeling: the non-mathematical part of the analytical-hypothetical style.
    b. *history* style: a *bona fide* standard view style.

Further characterization of these styles can be found in Winther (2012, 2013). The high-level and broad contrast between mathematical modeling and function styles is very close to the distinction between *formal* and *compositional* biology (Winther 2006c). The image we present here is one of potential collaboration and integration, rather than irreducible conflict, across styles.

Evidence that the debates are distinct

Cross-referencing between these literatures is remarkably scarce. In literature on the Wright–Fisher debate, there is effectively no reference to the units of selection debate (e.g., Biology: Coyne et al. 1997, 2000; Wade and Goodnight 1998; Goodnight and Wade 2000; Philosophy: Hodge 1992; Morrison 2000, 2006; Skipper 2002; Plutynski 2005). Reciprocally, the Units debate makes little reference to the complex genetic models found in the Wright–Fisher controversy (e.g., Sober and Wilson 1998; Keller 1999; Sterelny and Griffiths 1999; Lloyd 1988, 2000a, b, 2005a, b; Kerr and Godfrey-Smith 2002; Wilson 2003; Waters 2005; Okasha 2007). We provide further evidence for these claims in an online appendix.

One possible exception to the significant lack of cross-referencing is Wright (1980), who drew a distinction between the group selection criticized in the work of Maynard Smith, Williams and Dawkins, and group selection "for organismic advantage to individuals," which he had argued in favor of, and considered far more robust. This distinction is fundamental to the Units debate, since its acceptance permits the one-to-one mapping of selection level and 'adaptation for'—i.e., both occur at the level of the individual. Similarly, at least two of the leading participants in the group-selection debate, E. Sober and D. S. Wilson, discard Wright's work on these grounds. They explicitly distinguish Wright's 1945 model of altruism from his "more general 'shifting balance' theory of evolution." (Sober and Wilson 1998, 59)

After all, Wright's shifting balance theory "explain[ed] the evolution of individual traits, such as coat color in guinea pigs, rather than social traits such as altruism." (59–60) Perhaps because Sober and Wilson are interested in group-level traits, rather than in individual-level characteristics (potentially with "a complex genetic basis," 60) evolving through arguably hierarchical selection, they do not cite any of Wright's work besides his 1945. Interestingly, the functional focus on "adaptation for…" itself thus seems to be used as a justification within the Units of Selection debate for ignoring the Wright–Fisher controversy.

Wright's distinction between group selection for group advantage and group selection for complex individual adaptations is an important one in both biology and the philosophy of biology. Group selection for group advantage tends to imply opposition between levels of selection, especially given the adoption by participants of the Units debate of Williams' Occam's Razor (1966) favoring the lowest level of selection in explanations of adaptation., The necessary properties for an entity to function as a *unit* of selection have typically included properties, such as 'screening off,' (see below), implying the presumption of opposing selection levels. However, selection can also operate in the same direction at distinct levels (e.g., Otto and Hastings 1998).

In the following discussion, we do not wish to adjudicate among the conflicting perspectives within the two discourses. Instead, we wish to demonstrate that each of these literatures offers a distinct analysis both of multi-level selection and of the evolutionary process in general, while agreeing on the basic data, overarching mathematical framework, and pertinent problems and questions. An analysis of the two averaging strategies provides our entrance point into the assumptions and methods of the two literatures.

Article structure

In "The Wright–Fisher debate and complex genetics," and "The units of selection debate and the concept of "adaptation for …"," we present the fundamentals of each debate. In "Horizontal and vertical averaging strategies", we discuss two kinds of strategies of averaging gene effects. Comparing the implementation of these strategies across the two debates sheds light on similarities and differences in goals and methods of the Wright–Fisher and Units debates. "Style pluralism in evolutionary controversies" explores two distinct styles of explanation, modeling, and inference for evolutionary biology in general, and evolutionary genetics in particular: mathematical modeling versus function. We argue that evolutionary theory is pluralistic in its practices and that understanding evolutionary genetics (and negotiating its theoretical, conceptual, and methodological complexity) requires investigating its style pluralism.

The Wright–Fisher debate and complex genetics

This debate stems from strong disagreements between two seminal figures in twentieth century evolutionary theory, R. A. Fisher and S. Wright (e.g., Fisher 1918, 1937, 1941, 1953, 1958; Wright 1929, 1930, 1931, 1959, 1969, 1980; Provine 1986, Hodge 1992). Fisher's basic model of evolution stressed adaptive evolutionary change as driven by mutation and natural selection acting among individuals in large panmictic populations. He acknowledged and accounted for gene interactions (epistasis or G × G) and genotype-by-environment interactions (G × E), but found that they could be averaged over to obtain effects associated with single genes. Wright's basic model, the Shifting Balance Theory (SBT), was more complicated, with three phases: (1) random genetic drift within local populations (demes) moving demes across rugged adaptive landscapes; (2) mass Fisherian selection within demes, moving each up to the closest adaptive peak; and, (3) interdemic (group) selection, wherein demes at higher peaks send out relatively more migrants than demes at lower peaks, thereby shifting other demes to the gene combinations of the higher peaks (Wright 1977; Provine 1986; Wade 1996; Johnson 2008). Wright developed his theory to permit direct selection on genetic complexities, especially epistatic gene interactions (G × G). The combination of complex genetic systems and genetically subdivided populations results in multiple local fitness optima separated by valleys of lower fitness rather than a single Fisherian optimum.[1]

Fisher defined the average effect of a gene on fitness as an average over all gene interaction contexts: within an organism, across (related) organisms, and across environments. Variations in average effects from one gene to another defined his additive component of "*genetic* variance" which he claimed was "usually a considerable fraction" of the total genotypic variation (Fisher 1953, 271). Fisher saw averaging as a useful heuristic, but Wright disagreed: "Genetic variance was here [with respect to the "fundamental theorem of natural selection"] defined as merely the additive component, rather than the total variance due to heredity" (Wright 1959, 120). While Wright agreed that the additive variance determined the rate of change of local mean fitness, he considered the emigration of individuals from particular demes essential because it allowed unique allele combinations to spread. Wright saw the balance among the random and directional evolutionary forces (mutation, recombination, drift, mass intrademic [individual] selection, and interdemic selection among demes) as continually shifting.

The differences between Fisher and Wright can be further highlighted by analyzing their respective ways of understanding the components giving rise to individual phenotypes. In evolutionary genetics, the phenotype of individual i (Z_i) can be partitioned into contributions from genes in its own genome (G_i), contributions from genes in the genome of its social partners (G_si) and contributions from its environment (E_i).[2] In addition, each of these effects may interact giving rise to other influences on the phenotype, including interactions between genes in the

---

[1] Wright acknowledged that three dimensional landscapes are themselves a heuristic device and that there were many dimensions along which selection might act. More recent work (e.g., Gavrilets 1999) proposes that, in very high dimensional landscapes, all peaks are connected by ridges, so that the need for traversing fitness valleys by drift essentially disappears. Whether nature traverses valleys or ridges, however, remains an open question.

[2] There are additional genetic and environmental influences, not mentioned (e.g., Wolf et al. 2000).

same genome, such as classical epistasis (G_{i} x G_{j}); interactions between genes in the focal individual and those of its social partners, i.e., inter-genomic epistasis (G_{i} x G_{si}); and genotype-by-environment interactions (G_{i} x E_{i}). Thus, the hypothesized causal structure contains many terms and their interactions:

$$\begin{aligned} Z_{\mathrm{i}} = G_{\mathrm{i}} + G_{\mathrm{si}} + E_{\mathrm{i}} + (G_{\mathrm{i}} \times G_{\mathrm{j}}) + (G_{\mathrm{i}} \times G_{\mathrm{si}}) + (G_{\mathrm{i}} \times E_{\mathrm{i}}) \\ + \cdots (\text{higher order interactions}) \end{aligned} \tag{1}$$

There are an inordinately large number of higher order interactions because (1) there can be up to *N*th order classical epistasis [e.g., ($G_{\mathrm{i}} \times G_{\mathrm{j}} \times G_{\mathrm{k}}$)], inter-genomic epistasis [e.g., ($G_{\mathrm{i}} \times G_{\mathrm{si}} \times G_{\mathrm{sj}}$)] and genotype-by-environment interactions [e.g., ($G_{\mathrm{i}} \times G_{\mathrm{j}} \times E_{\mathrm{i}}$)], and (2) some of these include interactions among classical epistasis, inter-genomic epistasis, and genotype-by-environment interactions. Since each of these terms contributes orthogonally to Z_{i} and can be partitioned as a distinct variance component, we are not double counting. The averaging strategy over the *horizontal* plane, to be discussed below in "Horizontal and vertical averaging strategies", occurs when the effect of a focal gene is averaged across classical and inter-genomic epistasis as well as genotype-by-environment interactions. It is important to note, however, that the components of Z_{i} in Eq. (1), including G_{i} (the additive effect), change in magnitude with changes in gene frequency or in environment frequency (see below).

Phenotypic variation among individuals within a population can, thus, be characterized in terms of variations from one individual to the next in the components of Eq. (1). The terms G_{i} and E_{i} are central to arguments of local individual selection and individual adaptation. In contrast, the *contextual terms*, which include all the interactions, are most important for among-group selection because they mean that a gene's contribution (G_{i}) to the phenotype can be altered by context. Importantly, the rank order of direct allelic effects (G_{i}) on the individual phenotype can be changed by altering any of the contextual terms. That is, an allele at one locus can be "better" than another in one context, but "worse" than that other allele in a different context. Wright viewed random genetic drift operating in small local demes as one mechanism for creating variation among demes in genetic context. Estimating the sign and magnitude of the contextual terms requires different and more complicated experimental designs than estimating G_{i} and E_{i} (Demuth and Wade 2007a, b), and the estimated values of G_{i} and E_{i} can be misleading when the contextual terms are ignored. Fisher averaged over interactions to define a global effect of an allele independent of genetic background and environment, while Wright argued that much of evolution occurred in small populations, where specific local interaction effects predominated. Thus, contextual terms were relatively unimportant for Fisher, but critically significant for Wright. However, both investigated complex genetic evolutionary processes and agreed that modeling these processes was the aim of an evolutionary theory of individual adaptation.

Wright and Fisher also differed in their views on the most realistic and appropriate simplifying assumptions and idealizations to make in modeling nature. For Fisher, the large population size assumption (see Wade 1992) permitted his development of the very useful concept of gene effect; it mitigated random drift, but

Fisher considered that unimportant.[3] Similarly, in his island model, Wright idealized local population size, the reproductive behavior of individuals, and migration. These assumptions permitted him to derive the distribution of gene frequencies among demes. Real world variations in offspring numbers, sex ratio, etc., could be calibrated to the idealized theory via Wright's concept of *effective population size*.

A final way of emphasizing the differences between Fisher and Wright is by turning to a contemporary discussion of the "mean field assumption." (Bar-Yam 2000) According to Bar-Yam, the mean field assumption is "*…equivalent to allowing each of the components* [of a system] *to be placed in an environment which is an average over the possible environments formed by the other components of the system*" (Bar-Yam 2000, 279). Fisher alluded to such a set of assumptions in noting "that the fundamental theorem [of natural selection] proved above bears some remarkable resemblances to the second law of thermodynamics" (1958, 36–37). In assigning fitness effects to single genes as per Fisher, each gene is averaged in every possible genotypic, environmental and ecological context. As soon as populations are genetically subdivided, this assumption is violated because reproduction (local random mating) no longer mixes a gene into the genotypic combinations unique to other populations. Thus, a gene's effect becomes characteristic (at best) only of the 'local' background across which it is averaged. More importantly for Wright's theory, the effect of a gene on fitness can differ between different local backgrounds, even to the point of varying in sign.

Modern defenders of each theory appeal either to parsimony and explanatory power (Fisherians) or explanatory power and general unification (Wrightians) to justify the validity of their perspective. The most recent exchanges on these matters can be found in the journal *Evolution* (Coyne et al. 1997, 2000; Wade and Goodnight 1998; Goodnight and Wade 2000). Indeed, the energetic (even acrimonious) persistent disagreement of this exchange (see Feldman and Warfield 2010), as well as the close attention to details (e.g., differing interpretations of the causal structure of selection in cage selection of chickens, and in the case of cytoplasmic incompatibility, e.g., Winther 2006a, 219–220), strongly suggests that they are engaging with the same aim of characterizing the formal structure of the causes of hierarchical and adaptive evolution, and are not merely asking different, cross-cutting questions. In other words, they have the same concerns (e.g., the relative frequency of population structure and epistasis) about which they reasonably and genuinely disagree.

The units of selection debate and the concept of "adaptation for…"

One primary concern of the appropriate conceptual analysis of the functional concept *adaptation for …* over the last four decades is attribution of adaptation to different levels: the individual organism, the group, or both. A central question for

---

[3] For further discussion of Fisher's realism, see, e.g., Frank and Slatkin (1992). In interpreting Fisher's Fundamental Theorem of natural selection, they highlight "Fisher's ecological, holistic view, and the very reasonable interpretation of clutch size that follows from this view." (94).

this debate continues to be: “When a population evolves by natural selection, what, if anything, is the entity that does the adapting?” (Sober 1984, 204). A corollary of this question is its inverse: If we identify the entity (e.g., individual or group) that benefits by an adaptation, can we infer the nature of the historical causal process that gave rise to that adaptation? The recent triumph of a recursive, hierarchical, “transitional” picture of levels of biological organization (e.g., Maynard Smith and Szathmáry 2009) has both modified and intensified the interest in these questions.

Key participants of the Units of Selection controversy reacted to the work of the Chicago School of “Physiological Ecology” ca. 1930–1950, where Warder C. Allee, Alfred E. Emerson and Thomas Park were constructing a complex, hierarchical view of evolutionary ecology, and informally collaborating with Wright (see Wright 1959, 143–44 where he cites Wheeler, Allee and Emerson; Mitman 1992; Thomas Park, pers. comm.). For example, in the 1996 preface to his noted 1966 book, G.C. Williams recalls his time as a post-doc at the University of Chicago in the 1950s, and in particular his critical reaction to a presentation given by Emerson on the adaptive function of senescence for populations. However, the orthodox history traces this discourse to Williams (1966) and Dawkins (1976) and to their attacks on Wynne-Edwards (1962) (see Borrello 2003, 2010 for a sympathetic reading of Wynne-Edwards; Hamilton and Dimond 2012). Williams’ and Dawkins’ work from the 1960s and 1970s adopted an idealized rendition of the Fisherian genetical viewpoint (see Williams 1966, 57; Dawkins 1976; Wade 1980).

The publication of Wynne-Edwards (1962) and subsequent responses is considered the beginning of the modern group selection debates in biology. However, Wynne-Edwards was responding to Lack’s (1954) explanation for why animal populations do not seem to overexploit their available resources (Wynne-Edwards 1985). As one example, Lack argued that deferred maturity in birds could be explained by harsh environments, which place a strain on inexperienced individuals attempting to raise young (Lack 1954, 1966). In contrast, Wynne-Edwards explained the same phenomenon as a “homeostatic mechanism” of the population to prevent younger individuals from entering the breeding pool, such as through limited breeding sites or other forms of dominance interaction via “epidiectic displays,” until there were sufficient resources to support them (Wynne-Edwards 1962). While each of these explanations accounts for the existence of deferred maturity, G. C. Williams argues that Lack’s explanations provide a more parsimonious causal explanation referring to a lower level in the biological hierarchy.

In the “Scientific Study of Adaptation” (last chapter of 1966; see also Cassidy 1978), Williams argued that (1) attributions of adaptation required significant evidence, (2) adaptation should first be hypothesized as a lower level phenomenon, and (3) lower level selection was almost always explanatorily sufficient.

Williams distinguished between an “organic adaptation,” an adaptation at the organismal level, and a “biotic adaptation,” an adaptation at the biotic level (group-level or higher). Williams did not view aggregates of organic adaptations as sufficient to be considered biotic adaptation. For instance, while a herd of fleet deer could be called a fleet herd, this was not sufficient for considering fleetness a biotic adaptation. That designation required more, such as “coordinated teamwork for producing the effect, or mechanisms for producing group benefit by individual self-sacrifice" (1966, 262). Williams then argued, that "unless there are such things as biotic adaptations, there is no need for the theory of group selection" (1966, 103). He cited the social insects as one possible case of a biotic adaptation (1966, 197) but concluded "the apparent absence of comparable organization at any group of unrelated individuals is cogent evidence of the unimportance of biotic adaptation". (1966, 200). Thus, Williams articulated a stringent concept of group-level "adaptation for…", and found it to be extremely rare in nature. Consequently, group selection was also exceedingly rare, if present at all, in the natural world (see also Williams 1992).

Brandon's, and Sober and Wilson's, subsequent positions on hierarchical selection and hierarchical adaptation critique Williams' perspective. In his analysis of hierarchies of interactors and replicators, Brandon expands upon Hull's concepts and provides a definition of a level of selection in terms of "screening off" (1982, 1990; see Reichenbach 1956):

> Selection occurs at a given level if and only if (1) there is differential reproduction among the entities at that level; and (2) the adaptedness values of those entities screen off the adaptedness values of entities at every other level from reproductive values at the given level. (1985, 91)

On this view, adaptations have properties of "adaptedness" (roughly: the accuracy of fit with environmentally-given design problems) and these account for actual differences in reproductive success (see also Burian 1983, Krimbas 1984). Furthermore, the methodology of 'screening off'[4] reproductive values [roughly: realized[5] fitness] provides a way of simultaneously identifying the appropriate level of selection and the appropriate level of adaptation. Thus, the adaptedness of a trait is logically and causally prior to its having certain realized fitness values, but identifying the realized fitness value gives us empirical and epistemic access to determining the adaptedness of the trait.

As one consequence of his argument, Brandon holds (similarly to Williams) that group adaptation and group selection are tightly linked, and that (in contrast to Williams) adaptation can be correctly ascribed to interactors at a variety of levels. That is, "if we admit the possibility of selection occurring at levels other than that of the individual organism (and I don't see how anyone can at present deny that possibility), then we must admit the possibility of adaptations occurring at other levels" (1985, 93). Brandon is more optimistic than Williams about the frequency of group adaptations and group selection, insisting that "an explanatory adequate theory of adaptation requires a hierarchical and ecological theory of natural selection" (1990, 194).

Screening-off is a powerful tool for characterizing group selection and group adaptation, according to Brandon. First, consider cases where the mean fitnesses of individuals between groups differ but mean group fitnesses do not differ (i.e. MLS1,

---

[4] Brandon's characterization: "If A renders B statistically irrelevant with respect to outcome E but not vice versa, then A is a better causal explainer of E than is B" (1990, 83).

[5] As opposed to dispositional fitness.

Damuth and Heisler 1988). Because group structure itself is not under selection, such a scenario “cannot lead to the evolution of group adaptations and is not a case of group selection” (1990:116). These are cases where group structure merely affects individual fitness, and adaptedness values of individual entities screen off the values of higher-level entities. Thus, these cases should be differentiated from cases of genuine group selection. Second, Brandon discusses levels of adaptation in general, but does not provide an explicit analysis of what a group adaptation would be. Thus, it is difficult to know exactly which sorts of criteria for claims of group adaptation Brandon would endorse, other than a screening-off analysis.

Wilson and Sober (1989) and Sober and Wilson (1998, 2002, 2011) provide key defenses of group selection and group adaptation, and identify a clear causal relation between them: “When between-unit selection overwhelms within-unit selection, the unit itself becomes an organism in the formal sense of the word” (1989, 343). Indeed, group selection produces superorganisms: “a collection of single creatures that together possess the functional organization implicit in the formal definition of organism” (339). The essential attribute, the “hallmark” of an organism, is “functional organization” (339; see also 1998, 10–11). Note that Sober and Wilson use ‘organism’ in the same way that Brandon, following Hull, uses ‘individual.’ Each is employing a word identifying a higher-level functional unit, a supra-unit that is more than the sum of the units that comprise it.

In counteracting strongly individualist tendencies in the study of adaptation, Sober and Wilson agree with Williams’ functional definition of group “adaptation for… ”.[6] Moreover, they argue that “Williams’ Principle” (Sober and Wilson 2011) that an adaptation at a level requires selection at that same level, should serve as a common criterion for those in favor of and against group adaptation (Sober and Wilson 1998, 102). They disagree with Williams over the empirical frequency of group adaptation, which they believe to be relatively high in nature and to not be particularly susceptible to Occam’s razor and arguments from parsimony. After all, Sober and Wilson have less severe criteria for the application of the group adaptation concept: (1) group adaptation need not necessarily imply strong group integration, such as Williams’ criterion of “coordinated teamwork” since integration itself can be viewed as an adaptation, and (2) group adaptation is consistent with individual adaptation, e.g., “altruism is only one kind of group-level adaptation” (1998, 30–31). This also means that traits might result from both individual and group level selection, and thus be both individual and group adaptations.[7]

---

[6] The functions literature is central to the philosophy of biology. A detailed investigation of this literature would be necessary for understanding the function style in general and for comprehending, for instance, Sober and Wilson’s views on group functionality (e.g., Kitcher 1993; Allen et al. 1998; Ariew et al. 2002; Buller 1999). Standard analyses focus on the functions of individuals and their parts; interlocutors in this discourse have unfortunately not addressed group functionality in any detail. We believe that it would be fruitful to examine the concept of *group function*, especially as it relates to concepts of *group adaptation* and *design*.

[7] This could be legitimately interpreted as a concern with adaptive process as well as adaptive product. However, we wish to stress that adaptive process requires fleshing out the evolutionary ecological details, something the predictive and explanatory theory of evolutionary genetics arguably does in a much less abstracted way than the Units of Selection debates.

From this discussion of two important reactions to Williams (1966) from the Units of Selection debate literature, Brandon's and Sober and Wilson's, it should be clear that an important aim is to achieve clarity regarding the concept of higher-level "adaptation for…", its criteria of application, and its causal basis.[8]

Horizontal and vertical averaging strategies

Both the Wright–Fisher and Units of Selection debate make central reference to the strategy of averaging for inferring genetic effects, and for pinpointing the action of natural selection. There are defenders and critics of averaging within each of these debates. The forms of arguments used in the Wright–Fisher and Units of Selection controversies about averaging differ in part because the styles—mathematical modeling and function, respectively—differ across the two debates. Moreover, there are two forms of averaging:

1. *Horizontal,* i.e., averaging across (parts of) genomic context
2. *Vertical*, i.e., averaging across fitness ascriptions (of the same locus, or same set of loci) of units of selection at distinct hierarchical levels

There are thus three distinctions, and eight positions, to be outlined (Wright-Fisher/Units of Selection, critics/defenders, and horizontal/vertical averaging; see Table 1 below).

Below, we illustrate two basic pictures of averaging. Next, we characterize horizontal and vertical averaging. Finally, we explore the plurality of positions on averaging.

Dawkins' Oarsmen and Sober's causes

As an example of the averaging argument, consider Dawkins' well-known account of the Oxford oarsmen renders a vivid image of the averaging strategy. In this gedanken experiment, Dawkins asks us to consider how one would choose the best individual oarsmen from a large pool of candidates. He suggests that we should select the best by averaging the performances of each individual across a large number of random crew aggregations, and under different environmental conditions (1976, 40–41). Ranked performance can be assessed by grouping each oarsman with all possible partners, tabulating numbers of wins (or runner-up, etc.) for each oarsman, and then rank ordering them. Because of differences in external environment, each oarsman would also have to be tested under all external rowing conditions (40). But once we have averaged over all possible partners and environments, we will have an objective, ranked list of the best oarsmen. After all, "… a gene [oarsman] which is *consistently* on the losing side is not unlucky; it is a bad gene [oarsman]." (1976, 41) In "Horizontal averaging/defender/Units of

---

[8] At least one outstanding issue, inspired by Gould and Lewontin (1979), is the potential danger of what could be called *group adaptationism*.

Table 1 The topography of perspectives on averaging

	Defenders of the averaging strategy		Opponents of the averaging strategy	
	Wright–Fisher debate	Units of Selection debate	Wright–Fisher debate	Units of Selection debate
Horizontal Averaging	1. Fisher	2. Williams, Dawkins, Sterelny and Kitcher	3. Wright, Lewontin, Wade	4. Wimsatt
Vertical Averaging	5. Coyne, Barton and Turrelli	6. Williams, Dawkins, Sterelny and Kitcher	7. Wright, Wade	8. Wilson and Sober

See text for discussion

Selection", we discuss how Dawkins applies this averaging strategy to finding the best genes (actually: alleles).

As an example of the averaging fallacy argument, consider the penultimate section of Sober's *The Nature of Selection*:

> The strategy of averaging reduces to a parsimonious uniformity the variety of causal structures that a selection process may exhibit. But this uniformity arises from glossing over just the sorts of differences that the idea of a unit of selection ought to mark. (Sober 1993, 355)

As one of the main critics of averaging strategies, Sober has steadfastly argued that averaging destroys causal and mechanistic information about the selective process. It is trivially true that selective processes can be mathematically *redescribed* as a process happening on the gene level, but such retranslation fails to describe the actual selective causal processes (Sober and Lewontin 1982; Sober 1993; Sober and Wilson 1998) (and, moreover, only holds under certain formal assumptions e.g., Wade and Goodnight 1998; Welsh et al. 1988). In short, Sober is interested in functional concerns, and argues that these belie the averaging strategy.

Horizontal averaging and vertical averaging

Horizontal averaging is an averaging across genomic context. That is, the fitness effect of allelic substitution (e.g., A → a) is calculated by assessing the fitness of each genotype at a specified locus (e.g., AA, Aa, and aa) across all possible background genotypes, each weighted by its relative frequency in the population, and across all environments, each weighted by its relative frequency of encounter. Fitness effects at a particular locus are thus *ceteris paribus*.

Note that defenders of horizontal averaging can respond that there is always a "grand multi-dimensional mean" within which all combinations of oarsmen or genes (across all environments) are tried. In contrast critics of horizontal averaging appeal to a "localism"—neither natural selection nor the coach have the time or resources to test all possible combinations, and there is therefore no grand mean. Furthermore, local races do not include all combinations of oarsmen. Thus, the average rank might well be insufficient to predict (or deconstruct) race outcomes.

This contradicts the rationale for the averaging exercise, which is that while boats and their rowers win or lose *as a group*, it is the abstracted performance of the sum of the individual rowers that determines the outcome of races.

Vertical averaging is an averaging over the hierarchical fitness structure of a single locus. At each level, the response to selection equals the product of the selection differential and the heritability. In the simplest, one-locus case without interactions, the Price equation (Price 1970; Wade 1980 for an application to kin selection) allow one to partition the selection differential acting on a single gene into within and between group components, and Wright's one-locus F-statistics allow a parallel partitioning of the within and among group genetic variance.

Averaging over selection processes, whether simple to do or greatly complicated, eliminates the *causal* story of hierarchical selection. Genes with a positive marginal fitness effect, whether altruistic or not, increase in frequency. This sense of averaging over processes is different from the former averaging over fitness effects whenever there are gene interactions. Following terms in Joseph (1980), these background loci are not *ceteris paribus* here as they were in horizontal averaging, but *ceteris absentibus*. Averaging occurs vertically over different interactor levels of selection, rather than horizontally across background replicator genomic context. (Conversely, in averaging across genomic context, vertical averaging may or may not be done or even considered; when it is not, higher-level structure becomes *ceteris absentibus*.)[9]

In both cases of averaging, the type of averaging must also be carefully considered (e.g., geometric mean, arithmetic mean or harmonic mean), especially by defenders of averaging in evolutionary genetics such as Fisher and Coyne, Barton and Turelli. Care is required because use of the wrong average can be misleading. For example, in regard to population growth rates, Lewontin and Cohen (1969) illustrated that, for the same data, the arithmetic mean can be positive, suggesting a population increasing to infinite size, while the geometric mean may be less than 1, indicating a certainty of eventual extinction. Here, the arithmetic mean is a misleading heuristic. Similarly, Wright (1931) showed that, for populations whose size varies either in time or space, the harmonic mean and not the arithmetic mean is the appropriate average for describing the effects of random genetic drift.

Three distinctions, eight positions

In this section, we explore six of the eight perspectives resulting from the three distinctions that we have identified. More could be said about each position, but motivating intuitions will have to suffice. For cells 5 and 7, the interested reader is invited to consult the relevant primary literature as well as, e.g., Hodge (1992), Skipper (2002), and Winther (2006a). Note that in addition to defending or opposing an averaging strategy, a third option is always possible: remaining silent. For instance, opponents of horizontal averaging may remain silent about vertical averaging (e.g., Lewontin). For purposes of simplicity, we do not consider the third alternative here.

---

[9] For a perspective on averaging different from ours, see Okasha (2004, 2007).

Horizontal averaging/defender/Wright–Fisher

The founder of the horizontal averaging strategy in the Wright–Fisher debate is Fisher himself:

> the population used to determine [the] value [of "the *average excess* (in stature) associated with the gene substitution in question"] comprises, not merely the whole of a species in any one generation attaining maturity, but is conceived to contain all the genetic combinations possible, with frequencies appropriate to their actual probabilities of occurrence and survival, whatever these may be, and if the average is based upon the statures attained by all these genotypes in all possible environmental circumstances, with frequencies appropriate to the actual probabilities of encountering these circumstances. (1958, pp. 30–31)

Notice that Fisher is averaging across all *possible* genetic background, and across all *actual* environmental circumstances to obtain a multilocus, multi-environmental grand mean average. For such a broad averaging to be an effective strategy, a particular assumption must be met: the population of a species must be extremely large as well as randomly mating and thoroughly mixed across its entire range. Indeed, in a letter to Wright, Fisher wrote: "I believe that *N* must usually be the total population on the planet" (Provine 1986, 255).

Horizontal averaging/defender/Units of Selection

In his Oxford oarsman analogy, Dawkins is explicit that averaging is occurring across genomic context—i.e., is horizontal. In making the analogy explicit, he writes:

> The oarsmen are genes. The rivals for each seat in the boat are alleles potentially capable of occupying the same slot along the length of a chromosome. Rowing fast corresponds to building a body which is successful at surviving. The wind is the external environment. The pool of alternative candidates is the gene pool. As far as the survival of any one body is concerned, all its genes are in the same boat. … (1976, 40–41)

Just like we would have to test a particular oarsman *against* all its other possible competitors (sitting in the same place on the boat, e.g., left vs. right side, front or back), with respect to *all* background conditions (e.g., oarsmen sitting elsewhere in the boat, and weather), natural selection has to test a particular allele *against* all other possible alleles at that locus, with respect to *all* background conditions (especially alleles at *other* loci, i.e., other genes, and random strokes of luck). Averaging is performed across genomic context.

Horizontal averaging/opponent/Wright–Fisher

Critiques of horizontal averaging are found in Lewontin (1974) and Goodnight and Wade (2000). Lewontin (1974) argues that:

> Genes in populations do not exist in random combinations with other genes. The alleles at a locus are segregating in a context that includes a great deal of correlation with the segregation of other genes at nearby loci. The fitness at a single locus ripped from its interactive context is about as relevant to real problems of evolutionary genetics as the study of the psychology of individuals isolated from their social context is to an understanding, of man's sociopolitical evolution. In both cases context and interaction are not simply second-order effects to be superimposed on a primary monadic analysis. Context and interaction are of the essence. (318).

More specifically, and alluding to Lewontin's work on dimensional insufficiency for explaining and predicting gene frequency changes with only single-gene fitness measures (see also Wimsatt 1980), Goodnight and Wade (2000) write:

> Research into the genetic effects of epistasis has taken the approach of adding second order interaction terms to models in which alleles are assigned constant additive effects a priori. … We believe that this approach does not adequately address our view that the "additive effects" of genes are a property of the interactive system and *not* properties of individual genes. [Citation to last sentence of Lewontin 1974, p. 318, see above.] …In our view, the additive effect of a gene is strictly a statistical concept, a marginal value that summarizes the web of genetic interactions. It is not a free standing, independent property of the gene itself that can be treated in the theory like a constant with an assigned value. (319)

These passages emphasize the context-dependence of phenotypic/fitness values assigned to different alleles of single genes. Because populations are finite, mating is non-random, and time is limited, not all possible combinations can actually be tried. It is thus a fiction to confuse *all possible* combinatorics—and the in principle, tautologous averaging it justifies—with *only actually* combinatorics, and the local, context-dependence phenotypic/fitness value ascription that it implies. The horizontal averaging strategy is a reification of "all possible" combinations. It gives rise to misleading descriptions of the complex genetics of the processes of natural selection.

Horizontal averaging/opponent/Units of Selection

Wimsatt (2007, 1980) criticized genic selectionist models as inappropriate and inadequate computationally. Wimsatt (1980, 157) states: ''Illegitimate assumptions of context-independence are a frequent error in reductionist analyses''. For this reason, Wimsatt (1980) sees Williams' horizontal averaging leading to "a kind of genetic bookkeeping" rather than "promising a reductionistic theory of evolutionary change in terms of gene frequencies" (158). Moreover, functional concerns are crucial to Wimsatt, who writes "the most general maxim for those who study functionally organized systems is that we come to understand how things work by studying how, when, and where they break down" (2007, 22).

Vertical averaging/defender/Units of Selection

Sterelny and Kitcher (1988, 345) advocate averaging since it usefully employs "the strategy of abstracting from the thousand natural shocks that organisms in natural populations are heir to." Why abstract?

> 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, and, to this end, the *strategy of averaging*, which Sober decries, is entirely appropriate. (1988, 345, emphasis added)

Although they do not distinguish between the two forms of averaging here identified, they are taking issue with Sober, who explicitly argues against vertical averaging. Thus, we consider them to be implicitly arguing for vertical averaging. It is interesting to note that defenders of horizontal averaging often seem to take it for granted that such averaging gives vertical averaging for free, perhaps because global horizontal averaging seems to destroy any local group structure. Thus they believe that vertical averaging need not be addressed.

Sterelny and Kitcher employ the averaging strategy to emphasize the adaptive or maladaptive value of particular genes. For instance, they point to reading as a highly complex trait that involves many different genes, and likely environmental interactions as well. However, there are conditions such as dyslexia, which result from a mutation at a particular locus. In the sense that these mutations can be considered "genes for dyslexia" their counterparts can be considered "reading alleles" (1988, 351). Thus, different alleles of the same gene are either "good" or "wrong" (1988, 352). Discussing good or wrong alleles making a difference to phenotypic outcome locates function at the genetic level. Moreover, it is a clear example of employing a functional style of analysis (see also Kitcher 1993). Sterelny and Kitcher here focus on the notion of "adaptation for …". They worry about the genetic structure, and potential mechanisms, underlying adaptations qua products, in this case adaptations at the individual level.

Vertical averaging/opponent/Units of Selection

We saw Sober's opposition to averaging in "Dawkins' Oarsmen and Sober's causes" above. Sober and Wilson (1998) further critique the *vertical* averaging strategy:

> Another reason to reject the averaging approach is that it fails to identify the separate causal processes that contribute to the evolutionary outcome. When altruism evolves, there typically are two processes at work. Between-group selection favors the evolution of altruism; within-group selection favors the evolution of selfishness. …When this two-level process occurs in a population, an appropriate causal analysis should describe what is going on. The summary statement that the trait that evolved had the higher average fitness does not include any of these details. (32–33)

They defend a fact of the matter regarding the causal forces at play in the evolutionary process. The vertical averaging strategy incorrectly and unjustifiably distorts the representation of these selection forces.

Interestingly, towards the end of the book, Sober and Wilson also critique vertical averaging strategies for making presumptuous claims about *psychological* (and not just biological) altruism.This also shows their concern with *functional* matters, rather than with complex genetics and mathematical modeling. Indeed, Sober and Wilson's concerns with the appropriate causal-mechanistic story of hierarchical selection, and with the appropriate level of adaptation attribution (a judgment also potentially requiring historical, comparative analysis), are very much concerns of the functional style. Recall that functional modeling and explanation are bolstered by the styles of mechanism and history (again, see Winther 2012, 2013 for further discussion). Note that both Sterelny and Kitcher, and Sober and Wilson, are guided by conceptual analysis and explanatory strategies of the function style. They are not interested in working out complex mathematical modeling details of evolutionary genetics.

In "Horizontal and vertical averaging strategies", we have explored two kinds of averaging strategies and showed how they are operative (or blocked) in both the Wright–Fisher discourse, and the Units of Selection debate.

Style pluralism in evolutionary controversies

We hypothesize that at least part of the reason why these two controversies exist and interact relatively little is that they tend to be expressions of different *styles of modeling and explaining* the evolutionary process (e.g., Hacking 2002; Elwick 2007; Winther 2012, 2013). In particular, mathematical modeling style guides the Wright-Fisher debate, and the function style pertains to the Units of Selection discourse.[10] The latter itself requires two styles: history and mechanism (Winther 2012, 2013). These two styles are necessary but not sufficient for functional modeling, because concerns with "adaptation for…" must also be added.

Mathematical modeling aims at abstracting, idealizing, and generalizing a *mathematical model* for a particular set of objects or processes that express regularities and obey causal rules. The process of generating and using the model can be articulated in terms of five sequential activities: (1) *setting up*, (2) *manipulating*, (3) *explaining*, (4) *objectifying,* and (5) *pluralizing* (*SMEO-P* account, see Winther 2006a, b). Other philosophical models of mathematical modeling include Hughes' (1997) "Denotation, Demonstration, and Interpretation" or Bueno and Colyvan's (2011) "Immersion, Derivation, and Interpretation" pictures. The bulk of work in the Wright-Fisher debate concerns building mathematical models, ideal and complex, with which we can represent and intervene in the world.

In contrast, the mechanism style takes a functional system and breaks it down in order to understand how it works. Of which (types of) parts does the system consist?

---

[10] There are some exceptions to this pattern. For instance, Simpson's Paradox as used in the Units debate is mathematical, but the mathematics is used to make a conceptual point.

How do these parts behave, and what do they cause each other to do? What are the basic theoretical principles governing the parts, as well as the system as a whole? The mechanism style searches for and constitutes mechanisms using four overarching strategies: (1) *analysis*, (2) *physicochemical (PC) reduction*, (3) *causal surgery*, and (4) *mechanism transplantation* (see Winther 2013). Moreover, the history style in the biological sciences aims to present the narrative or biography of a part, which places it in its organizational and causal whole; this biography is justified by a phylogeny (Winther 2011). The Units of Selection discourse appeals to both these styles, in the context of the function style, when ascribing adaptation to the individual or the group, or both, *and* when suggesting a plausible historical narrative for the selection processes giving rise to the adaptation. Of course, it is possible to engage in the mechanism or history style without being concerned with function. This, however, is not the case for those interested in units of selection.

Of general interest is that evolutionary biology is a pluralistic explanatory and predictive enterprise and we must understand its style pluralism. Such pluralism, and its negotiation and potential integration, is also evident in research on complex systems (e.g., Mitchell 2003; Wimsatt 2007), and evolutionary developmental biology (e.g., Winther 2013).

Importantly, a single style does not fully guide one research domain (e.g., Wright-Fisher controversy), nor are styles mutually exclusive. They can intertwine within domains. Consider an example by way of conclusion. One of us (MJW) recounts how the faculty of Evolution and Ecology at the University of Chicago employed mathematical modeling, whereas faculty at the Field Museum was interested in functional inference. The pressure of DNA sequence data today is rearranging that picture. The former faculty is becoming “genetic museum faculty,” replacing experiments sensu stricto with sequence analysis, and adopting functional inference even as the Field Museum sets up facilities for storing DNA sequences from its specimens. Material practices can indeed change the application and nature of a style, as well as style relations.

Acknowledgments Peter Godfrey-Smith, MJS Hodge, Elisabeth Lloyd, Lucas McGranahan, Fabrizzio McManus Guerrero, Amir Najmi, Elliott Sober, Kim Sterelny, and two anonymous reviewers, provided useful feedback on earlier drafts. RGW was supported in part from a Faculty Research Grant from the Academic Senate Committe on Research, University of California, Santa Cruz, and by the Center for Models of Life, Niels Bohr Insitute, and the Center for the Philosophy of Nature and Science Studies, Copenhagen University; MJW thanks support from the NIH grant 5R01GM65414–4; CCD appreciates support from an NSF Graduate Research Fellowship.

References

Allen C, Bekoff M, Lauder G (eds) (1998) Nature’s purposes: analyses of function and design in biology. MIT Press, Cambridge

Ariew A, Cummins R, Perlman M (eds) (2002) Functions: new essays in the philosophy of psychology and biology. Oxford University Press, Oxford

Bar-Yam Y (2000) Formalizing the gene centered view of evolution. Adv Complex Syst 2:277–281. Url: http://arxiv.org/abs/physics/0002016

Borrello M (2003) Synthesis and selection: Wynne Edwards challenge to David Lack. J Hist Biol 36:531–566

Borrello Mark (2010) Evolutionary restraints. The contentious history of group selection. University of Chicago Press, Chicago

Brandon RN (1982) The levels of selection. PSA 1982(1):315–323

Brandon RN (1985) Adaptation explanations: are adaptations for the good of replicators or interactors? In: Depew DJ, Weber BH (eds) Evolution at a crossroads: the new biology and the new philosophy of science. MIT Press, Cambridge, pp 81–96

Brandon RN (1990) Adaptation and environment. Princeton University Press, Princeton

Bueno O, Colyvan M (2011) An inferential conception of the application of mathematics. Noûs 45(2):345–374

Buller D (ed) (1999) Function, selection and design. SUNY Press, Albany

Burian RM (1983) Adaptation. In: Grene M (ed) Dimensions of Darwinism. Themes and counterthemes in twentieth-century evolutionary theory. Cambridge University Press, Cambridge, pp 287–314

Cassidy J (1978) Philosophical aspect of the group selection controversy. Philos Sci 45:575–594

Coyne JA, Barton NH, Turelli M (1997) Perspective: a critique of Sewall Wright's shifting balance theory of evolution. Evolution 51:643–671

Coyne JA, Barton NH, Turelli M (2000) Is Wright's shifting balance process important in evolution? Evolution 54:306–317

Crombie AC (1994) Styles of scientific thinking in the European tradition. 3 vols. Duckworth, London

Damuth J, Heisler IL (1988) Alternative formulation of multilevel selection. Biol Philos 3:407–430

Dawkins R (1976) The selfish gene. Oxford University Press, Oxford

Demuth J, Wade MJ (2007a) Population differentiation in the beetle, *Tribolium castaneum.* I. Genetic architecture. Evolution 61:494–509

Demuth J, Wade MJ (2007b) Population differentiation in the beetle, *Tribolium castaneum.* II: Haldane's Rule and incipient speciation. Evolution 61:694–699

Elwick J (2007) Styles of reasoning in the British life sciences: shared assumptions, 1820–1858. Pickering & Chatto, London

Feldman R, Warfield TA (2010) Disagreement. Oxford University Press, New York

Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb 3:399–433

Fisher RA (1937) The design of experiments. Oliver & Boyd, Edinburgh

Fisher RA (1941) Average excess and average effect of a gene substitution. Ann Eugen 11:53–63

Fisher RA (1953) Population genetics. The Croonian lecture. Proc R Soc B 141:510–523

Fisher RA (1958) The genetical theory of natural selection. Dover Publications, Inc., New York (2nd edition of 1930 Oxford University Press 1st edition)

Frank SA, Slatkin M (1992) Fisher's fundamental theorem of natural selection. Trends Ecol Evol 7(3):92–95

Gavrilets S (1999) A dynamical theory of speciation on holey adaptive landscapes. Am Nat 154(1):1–22

Goodnight CJ, Wade MJ (2000) The ongoing synthesis: a reply to Coyne, Barton, and Turelli. Evolution 54:317–324

Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. Proc R Soc Lond B 205:581–598

Gould SJ, Vrba ES (1982) Exaptation—a missing term in the science of form. Paleobiology 8:4–15

Hacking I (2002) Historical ontology. Harvard University Press, Cambridge

Hacking I (2009) Scientific reason. National Taiwan University Press, Taipei

Hamilton A, Dimond CC (2012) Groups, individuals, and evolutionary restraints: the making of the contemporary debate over group selection. Biol Philos 27(2):299–312

Hammerstein P (1996) Darwinian adaptation, population genetics and the streetcar theory of evolution. J Math Biol 34:511–532

Hempel GC (1994) The logic of functional analysis (1959). Reprinted in Martin M, McIntyre LC. Readings in the philosophy of social science. MIT Press, Cambridge, pp 349–375

Hodge MJS (1992) Biology and philosophy (including ideology): a study of Fisher and Wright. In: Sarkar S (ed) The founders of evolutionary genetics. Kluwer, Dordrecht, pp 231–293

Hughes RIG (1997) Models and representation. Philos Sci 64:S325–S336

Johnson N (2008) Sewall Wright and the development of shifting balance theory. Nature Education 1(1). http://www.nature.com/scitable/topicpage/sewall-wright-and-the-development-of-shifting-30508

Joseph G (1980) The many sciences and the one world. J Philos 77:773–791

Keller L (1999) Levels of selection in evolution. Princeton University Press, Princeton

Kellert SH, Longino HE, Waters CK (eds) (2006) Scientific pluralism (Minnesota studies in the philosophy of science). University of Minnesota Press, Minneapolis

Kerr B, Godfrey-Smith P (2002) Individualist and multi-level perspectives on selection in structured populations. Biol Philos 17:477–517

Kitcher P (1993) Function and design. Midwest Stud Philos XVIII:379–397

Krimbas CB (1984) On adaptation, Neo-Darwinian tautology, and population fitness. Evol Biol 17:1–57

Kwa C (2011) Styles of knowing. A new history of science from ancient times to the present. University of Pittsburgh Press, Pittsburgh

Lack D (1954) The natural regulation of animal numbers. Oxford University Press, Oxford

Lack D (1966) Population studies of birds. Clarendon Press, Oxford

Lewontin RC (1974) The genetic basis of evolutionary change. Columbia University Press, New York

Lewontin RC, Cohen D (1969) On population growth in a randomly varying environment. Proc Natl Acad Sci USA 62(4):1056–1060

Lloyd EA (1988) The structure and confirmation of evolutionary theory. Greenwood Press, Westport

Lloyd EA (2000a) Groups on groups. Biol Philos 15:389–401

Lloyd EA (2000b) Units and levels of selection. An anatomy of the units of selection debates. In: Singh R, Krimbas C, Paul D, Beatty J (eds) Thinking about evolution: historical, philosophical and political perspectives. Cambridge University Press, Cambridge, pp 267–291

Lloyd EA (2005a) Why the gene will not return. Philos Sci 72:287–310

Lloyd EA (2005b) Units and levels of selection. In: Zalta EN (ed) The Stanford encyclopedia of philosophy (Fall 2005 edition). Url:http://plato.stanford.edu/archives/fall2005/entries/selection-units/

Maynard Smith J, Szathmáry E (2009) The origins of life. From the birth of life to the origins of language. Oxford University Press, New York

Mitchell S (2003) Biological complexity and integrative pluralism. Cambridge University Press, Cambridge

Mitman G (1992) The state of nature ecology, community, and American social thought, 1900–1950. University of Chicago Press, Chicago

Morrison M (2000) Unifying scientific theories: physical concepts and mathematical structures. Cambridge University Press, Cambridge

Morrison M (2006) Unification, explanation and explaining unity: the Fisher–Wright controversy. Br J Philos Sci 57:233–245

Okasha S (2004) The "averaging fallacy" and the levels of selection. Biol Philos 19:167–184

Okasha S (2007) Evolution and the levels of selection. Oxford University Press, Oxford

Otto SP, Hastings IM (1998) Mutation and selection within an individual. Genetica 102(103):507–524

Pickstone JV (2001) Ways of knowing. A new history of science, technology and medicine. University of Chicago Press, Chicago

Plutynski A (2005) Parsimony and the Fisher–Wright debate. Biol Philos 20:697–713

Price GR (1970) Selection and covariance. Nature 227:520–521

Provine W (1986) Sewall Wright and evolutionary biology. University of Chicago Press, Chicago

Reichenbach H (1956) The direction of time. University of California Press, Berkeley

Schwartz J (2002) Population genetics and sociobiology. Conflicting views of evolution. Perspect Biol Med 45:224–240

Skipper RJ (2002) The persistence of the RA Fisher—Sewall Wright controversy. Biol Philos 17:341–367

Sober E (1984) Sets, species, and evolution: comments on Philip Kitcher's "species". Philos Sci 51:334–341

Sober E (1993) The nature of selection. Evolutionary theory in philosophical focus. University of Chicago Press, Chicago [First Published MIT Press, 1984]

Sober E, Lewontin RC (1982) Artifact, cause and genic selection. Philos Sci 49:157–180

Sober E, Wilson DS (1998) Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Cambridge

Sober E, Wilson DS (2002) Perspectives and Parameterizations. Biol Philos 17:529–537

Sober E, Wilson DS (2011) Adaptation and natural selection revisited. J Evol Biol 24:462–468

Sterelny K, Griffiths P (1999) Sex and death. An introduction to philosophy of biology. University of Chicago Press, Chicago

Sterelny K, Kitcher P (1988) The return of the gene. J Philos 85:339–361

Uyenoyama MK, Feldman MW (1980) Evolution of altruism under group selection in large and small populations in fluctuating environments. Theor Popul Biol 17:380–414

Wade MJ (1980) An experimental study of kin selection. Evolution 34(5):844–855

Wade MJ (1992) Sewall Wright: gene interaction and the shifting balance theory. In: Antonovics J, Futyuma D (eds) Oxford surveys of evolutionary biology VI. Oxford University Press, NY, pp 35–62

Wade MJ (1996) Adaptation in subdivided populations: kin selection and interdemic selection. In: Rose MR, Lauder G (eds) Evolutionary biology and adaptation. Sinauer Associates, Sunderland, pp 381–405

Wade MJ, Goodnight CJ (1998) Perspective: the theories of Fisher and Wright in the context of metapopulations. When nature does many small experiments. Evolution 52:1537–1553

Waters CK (2005) Why genic and multilevel selection theories are here to stay. Philos Sci 72:311–333

Welsh AH, Peterson AT, Altmann AS (1988) The fallacy of averages. Am Nat 132:277–288

Williams GC (1966) Adaptation and natural selection. Princeton University Press, Princeton

Williams GC (1992) Natural selection: domains, levels, and challenges. Oxford University Press, Oxford

Wilson RA (2003) Pluralism, entwinement and the levels of selection. Philos Sci 70:531–552

Wilson DS, Sober E (1989) Reviving the Superorganism. J Theor Biol 136:337–356

Wimsatt WC (1980) Reductionist research strategies and their biases in the units of selection controversy. In: Nickles T (ed) Scientific discovery: case studies. Reidel, Dordrecht, pp 213–259

Wimsatt WC (2007) Re-engineering philosophy for limited beings piecewise approximations to reality. Harvard University Press, Cambridge

Winther RG (2006a) Fisherian and Wrightian perspectives in evolutionary genetics and model-mediated imposition of theoretical assumptions. J Theor Biol 240:218–232

Winther RG (2006b) On the dangers of making scientific models ontologically independent: taking Richard Levins' warnings seriously. Biol Philos 21:703–724

Winther RG (2006c) Parts and theories in compositional biology. Biol Philos 21:471–499

Winther RG (2011) Part–whole science. Synthese 178:397–427

Winther RG (2012) Interweaving categories: styles, paradigms, and models. Stud Hist Philos Sci A. Accessed Online 1 July 2012. http://philpapers.org/rec/WINICS

Winther RG (2013) Evo-devo as a trading zone. In: Love A (ed) Conceptual change in biology: scientific and philosophical perspectives on evolution and development, Boston studies in the philosophy of science. Springer Verlag, Dordrecht. Accessed Online 1 July 2012. http://philpapers.org/rec/WINEAA-4

Wolf JB, Brodie EB III, Wade MJ (eds) (2000) Epistasis and the evolutionary process. Oxford University Press, Oxford

Wright S (1929) Evolution in a Mendelian population. Anat Rec 44:287

Wright S (1930) Review of genetical theory of natural selection by RA Fisher. J Hered 21:349–356

Wright S (1931) Evolution in Mendelian populations. Genetics 10:97–159

Wright S (1945) Tempo and mode in evolution: a critical review. Ecology 26:415–419

Wright S (1959) Physiological genetics, ecology of populations, and natural selection. Perspect Biol Med 3:107–151

Wright S (1969) Evolution and the genetics of populations. The theory of gene frequencies, vol 2. University of Chicago Press, Chicago

Wright S (1977) Evolution and the genetics of populations. Experimental results and evolutionary deductions, vol 3. University of Chicago Press, Chicago

Wright S (1980) Genic and organismic selection. Evolution 34:825–843

Wynne-Edwards VC (1962) Animal dispersion in relation to social behavior. Oliver and Boyd, Edinburgh

Wynne-Edwards VC (1985) Backstage and upstage with "animal dispersion". In: Dewsbury D (ed) Studying animal behavior: autobiographies of the founders. University of Chicago Press, Chicago