Selfish Genes - The Eunuchs of Selection

Robin D. Knight & Paul E. Griffiths[1]

ABSTRACT

The idea that phenotypic evolution proceeds by the selection of genes relies on equivocating between different definitions of gene and systematically ignoring the type/token distinction. The Evolutionary Gene Concept (EGC) favoured by gene selectionists identifies DNA segments that have no nomological effects on their own reproductive success. Richard Dawkins's notion of an 'active germ line replicator' is an unsuccessful attempt to remedy this defect. The Classical Molecular Gene Concept (CMG) identifies DNA segments that are not typically in competition with one another. The key to rigorously representing gene-selection is to use the principles of modern systematics to specify type-identity conditions for genes. 'Gene selection' in various senses can then be defined, but none of these senses vindicates the gene selectionist vision.

[1] An earlier version of this paper was delivered at the International Society for History, Philosophy and Social Studies of Biology in July 1997. We are grateful for clarifying comments from other partipants in the developmental systems symposia, and from the audience. Particular thanks to Robert Perlman for additional examples.

1. Gene Concepts: An Embarrassment of Riches?

It is natural to suppose that the current popularity of gene selectionism, the view that the diversity of life is entirely explained by selection between genes, is somehow connected to the 'molecular revolution' in modern biology. Works in popular science and the philosophy of biology contain both fascinating accounts of molecular mechanisms involving 'genes' and accounts of ruthless competition between 'selfish genes'. Contrary to the impression conveyed to readers, the word 'gene' has completely different meanings in these two contexts. Gene selectionism requires that both senses of 'gene' pick out the same DNA sequences. Only then can phenotypic evolution be explained by the selection of the gene sequences involved in the development of phenotypes. But in reality the two senses of 'gene' coincide only occasionally. The genes of molecular biology are not typically in competition with one another for representation in future generations. The 'selfish genes' defined by gene selectionists typically have no systematic effect on their own replication. They are, to coin a phrase, the eunuchs of selection.

There is a tradition of related criticisms of gene selectionism to which Ernst Mayr, Robert Brandon, Elliott Sober and Richard Lewontin have all contributed (Mayr 1963; Brandon 1982; Sober and Lewontin 1982). The general thrust of these criticisms is that single genes have no consistent, nomological effect on fitness. The same gene can be fitness-enhancing in one context and fitness-reducing in another. William C. Wimsatt makes the point in a slightly different way. The fact that an allelic fitness value can be calculated post-facto from the actual fate of allele copies does not imply that the allele has a fitness value which could be used in advance of the fact to calculate its fate

((Wimsatt 1980 p229); see section 4 below). There is a well-known general reply to such criticisms: genes do have nomological fitness effects once their context is specified. Since the fitness of any unit of selection is necessarily relative to context, this reply does not weaken the explanatory power of gene selectionism (Sterelny and Kitcher 1988). The parallel reply to Wimsatt is that we could calculate the fate of an allele in advance if we knew the contexts in which it would find itself. Both objections and reply have real substance, as we explore below, but the debate between them has been fundamentally confused by a lack of attention to the identity conditions of genes.

A selective explanation explains the existence of entities of *a particular type* by showing that they reproduce more frequently than *other types.* The idea that genes are the units of selection thus presupposes an account of when two specimens of DNA are *the same gene*. This question, in turn, is ambiguous: it could be a question about gene tokens or about gene types. In order to make rigorous sense of gene selection, we need to know both the identity condition for a gene token (the equivalent of knowing what individual organisms are for a paradigmatic case of natural selection) and the identity conditions for gene types (the equivalent of knowing the traits of organisms in virtue of which they are selected). Obviously, Kettlewell could not have conducted his study of the evolution of melanism without knowing where one moth ended and the next began or without seeing dark-winged and light-winged as types that might be subject to a common selection process (Kettlewell 1973). Similarly, nothing coherent can be said about gene selection without an account of gene tokens and gene types (Biologists call a gene token a gene copy. A gene type is what a gene copy is a copy of.)

Modern biology contains many sub disciplines with distinct and often conflicting notions of what it is to be the same gene. We examine two of these concepts, the Evolutionary Gene Concept (EGC) espoused by George C. Williams and Richard Dawkins, and the Classical Molecular Gene Concept (CMG) identified by Eva Neumann-Held as the status quo in molecular biology (Neumann-Held 1997). When the gene types and tokens identified by these two concepts are contrasted, it turns out that the selfish genes identified by the EGC are typically not functional, and that the functional genes identified by the CMG concept are typically not selfish. These conclusions have dire implications for gene selectionism as a general account of evolution.

2. The Evolutionary Gene Concept (EGC)

The EGC is a version of the Mendelian gene concept that has been purged of some of the commitments that were problematic in the light of early molecular biology. A gene is 'that which segregates and recombines with appreciable frequency' (Williams 1966):24. The main virtue of the evolutionary gene from the point of view of the gene selectionist is that the generalisations of population genetics and evolutionary game theory can be applied to it. The proportion of different evolutionary genes in the next generation is a function of the average fitness of those genes in the last generation. Williams moves even further from molecular genetics in his later work, defining genes as units of pure information: 'DNA is the medium, not the message. A gene is not a DNA molecule; it is the transcribable information coded by the molecule. …the gene is a packet of information, not an object' (Williams 1992): 11. We do not consider the implications of this

definition here. Most gene selectionists who use the EGC maintain their allegiance to DNA: 'Genes are just DNA: they are dumb molecules.' (Dawkins 1995):221.

The EGC yields very liberal identity conditions for gene tokens: any token of DNA that is located on a chromosome in a sexual organism (so as to be able to segregate and recombine) and is reasonably short is a gene token.

The EGC yields a two-part identity condition for each gene type. Two gene tokens must have the same sequence and equivalent chromosomal locations to be instances of 'the same' gene. The same sequence condition is often relaxed to allow similar sequences with indistinguishable effect to count as the same gene type. The second, equivalent locus, condition is problematic for two reasons. First, loci are not static over evolutionary time. Although chromosomes are stable in inbred strains for the duration of a grant period, karyotypic change (change in the structure and number of chromosomes) can be quite profound even in recently separated species. Second, defining genes by their location makes gene duplication or gene transposition produce different genes. This blocks the gene selectionist intuition that duplication and transposition are *reproduction strategies* for genes. For these reasons, the second condition is often relaxed so that an evolutionary gene is of the same type as any sufficiently similar sequence elsewhere in the genome: 'Intragenomic selection of selfish DNA [deals with] the spreadability of certain kinds of DNA to *different* loci or the creation of new loci' ((Dawkins 1982):161, italics in original). This relaxed EGC seems to be the gene concept

characteristic of hard-line gene selectionism—the view that genes are the *only* real units of selection[2].

3. The Molecular Gene Concept

The Classical Molecular Gene (CMG) is a DNA sequence from which a particular molecular product is expressed. A gene token is therefore a token coding sequence together with the adjacent regulatory apparatus needed for its expression. This is the molecular gene to which optimists about the reduction of Mendelian to molecular genetics claim to reduce Mendelian genes (Waters 1990; Waters 1994). Although Eva Neumann-Held identifies the CMG as the primary gene concept of molecular biology to date, she questions its viability in the light of the recent developments ((Neumann-Held 1997), see also (Sarkar 1996)). The simplest problem is that many alternative gene products are produced from the same coding sequence and the 'regulatory apparatus' that must be added to specify a single molecular product includes many factors in the genomic and cellular context. This complicates the definition of a gene token, since it is unclear how much of this apparatus to include in the gene. It also complicates the definition of gene types, since whether two sequences are the same gene will depend how far down the chain of gene expression we look. Two sequences may have the same primary transcript but different transcripts after splicing or editing and hence different final products. Even a staunch defender of the CMG like Kenneth C. Waters admits that if genes are identified by their products, then which sequences are regarded as the same gene will change as molecular

---

[2] To many readers 'units of selection' will suggest the question 'Do they mean replicators, or interactors?'. We do not use the replicator/interactor distinction because it is not a well-drawn distinction. See: (Griesemer 1992; Griffiths and Gray 1994a; Griffiths and Gray 1994b; Griesemer Forthcoming). Those addicted to the distinction can read this paper as a discussion of whether genes can be 'active germ line replicators' (Dawkins 1982).

biologists move their attention from one stage of post-transcriptional processing to another (Waters 1990). One way to read Waters's proposal is that each CMG type has a particular norm of reaction across a range of genomic and cellular contexts. If two sequences consistently produce the same products when placed in common contexts, they are the same gene. We might call this the C(ontemporary) Molecular Gene concept. However, we are not sure that this is either a correct interpretation of Waters or a fully adequate solution.

These problems need not centrally concern us in this paper, since they concern the precise identity conditions for CMG's - the exact range of regulatory machinery that is part of the gene token and the conditions under which two token coding sequences are exactly the same type of CMG. These are both questions on which the actual practice of current molecular biology is relatively imprecise. Our concern will be with the broader categories of CMG with which molecular biology is usually concerned: the H4 histone gene found from humans to peas or IGF-2 (insulin-like growth factor -2) found from mice to horses. As we explain below, these concepts of 'the same' molecular gene are produced by the principles used elsewhere in biology to classify organisms into species or parts of organisms into anatomical categories. These principles include sheer physical similarity (in this case sequence similarity), shared ancestry, and shared function. The question that concerns us is whether the types identified by these principles have the key evolutionary property that variants on a type are in competition with one another for representation in future generations. In other words, are the genes identified by molecular biologists candidates for gene selection?

4. Two Constraints on the Units of Gene Selection

There are two constraints on any meaningful unit of gene selection (S-gene):

1. S-gene types must have some nomological effects in virtue of which they can be selected.

2. There must be in competition between tokens of an S-gene type, so that variants may be selected for or against.

Constraint 1 requires that the fitness value of a gene be due to that gene's causal powers affecting its reproduction rate. It excludes any fitness value that is a mere post facto average of what happened to the gene in the events with which it was historically associated. It is not enough that we can say copies of this gene *were* reproduced, we must be able to say that copies of such genes *would* have been or *will* be reproduced. Only a nomological effect, with counterfactual force, can figure in an *explanation* of evolutionary change. Gene selection is meant to be a causal process, not a mere coincidence.

The most fatuous form of gene selectionism simply averages the actual reproductive success of the tokens of some gene type, labels this figure an allelic fitness coefficient, and claims that this coefficient explains the current prevalence of the gene. The well-known analogy drawn by Dawkins between the genes in an organism and the crew of a racing rowing VIII encourages this interpretation, as does the accompanying slogan 'a gene which is *consistently* on the losing side is not unlucky; it is a bad gene.' ((Dawkins 1976):41, italics in original). *Pace* Dawkins, genes can be consistently unlucky and so absurd consequences follow immediately from this interpretation of gene

selectionism. There are innumerable non-coding stretches of DNA that are characteristically found in unsuccessful clades of organisms. These stretches of DNA have no biological effect whatever, and so cannot be 'worse genes' than similar non-coding stretches in successful species. The introns[3] in the glycogenin of the Dodo were not 'worse genes' than those in the corresponding sequence in cockroaches, yet only the latter survive. Ascribing them superior fitness is as absurd as ascribing superior fitness to the three surnames that survive today from those of the original settlers of tiny Pitcairn Island. Only a hard-line selfish genealogist would claim that 'a surname which is consistently on the losing side is not unlucky: it is a bad surname'.

Another delightful consequence of taking post-facto average fitnesses too seriously is that there are precisely four genes – the four DNA nucleotides A, T, G, and C – whose relative allelic fitnesses exhaustively explain the composition of all the DNA in existence. Each token of A, T, G or C at a particular point in a particular DNA molecule has occurred in an environment of two other adjacent bases, making up a potential codon (unless it occurs at the end of the DNA molecule and is hence in a non-coding region). Each of these potential codons has been part of a string of codons, and some of these strings have been functional classical molecular genes. For each individual token of a base, any of the three other bases could be substituted. In some cases the result would increase the fitness of the organism in which the change occurs, in others decrease it; in the vast majority of cases it would have no effect on fitness whatsoever. Despite this

[3] Introns are parts of genes whose corresponding sections in the RNA transcript are cut out and discarded before a protein is assembled using the transcript as a template.

variety of effects, we can derive an average fitness for a base such as thymine by averaging the fitnesses of the actual organisms in which each thymine token has occurred. Since the fitnesses of these organisms determine the number of copies of that thymine token that will be passed on to the next generation, the average fitness figure can be used to predict the proportions of bases in the next generation. Four such average fitnesses, for A, T, G and C, will 'explain' the composition of all the DNA in the world. Dawkins discusses the 'selfish-nucleotide' theory in The Extended Phenotype (Dawkins 1982):90. His attempt to avoid this *reductio ad absurdum* is examined in the next section.

5. Meeting the First Constraint

The constraint that S-genes must have nomological effects can be met by setting correct identity conditions for gene tokens. The genic equivalents of individual organisms must be units of molecular function, just as the units of organismic competition must be organisms and not the hind-quarter of a zebra plus the adjacent head of a lion. That meaningless organismic unit is the equivalent of Dawkins's pure evolutionary gene: 'When I said 'arbitrarily chosen portion of chromosome', I really meant arbitrary. The twenty-six codons that I chose might well span the border between two cistrons[4]. The sequence still potentially fits the definition of a replicator, it is still possible to think of it as having alleles…' (Dawkins 1982:87).

Dawkins attempts to meet the first constraint and avoid the selfish-nucleotide reductio in *The Extended Phenotype* (Dawkins 1982). He distinguishes *active* replicators 'whose nature has some influence over [their] probability of being

[4] Cistrons are CMG's in prokaryotes.

copied' (1982: 83) from merely passive replicators. But Dawkins's active replicators need not be units of molecular function. Just after the requirement that replicators be causally active we find the quotation above, which advocates arbitrary portions of chromosome! Although Dawkins accepts that CMGs (his 'cistrons') are good examples of active replicators (1982: 91-92), he is determined not to define the evolutionary gene ('optimon') in terms of molecular function. His final definition requires only that the arbitrary segments allowed by the relaxed evolutionary gene concept be fairly long: 'Fragments of DNA qualify as active germ line replicators. … they must not be defined too small if they are to be usefully regarded as active.' (1982: 95).

A careful reading of *The Extended Phenotype* reveals that Dawkins has not really come to terms with the lesson of the 'selfish nucleotide' reductio. The lesson is that an evolutionary explanation, like any scientific explanation, must have counterfactual force. Not every correlation is explanatory. If men born under the star-sign Leo had more sexual partners than average, we could appeal to the 'attractiveness of Leos' to explain this. The 'attraction coefficient' of Leos is obtained by averaging the attractiveness of individual Leos. The sexual success of Leos as a group can be explained by their attraction coefficient. Yet in reality the correlation must be either a fluke or the result of that date of birth correlating with some non-astrological property (such as being the oldest people in a cohort in the education system). The attraction coefficient of Leos is not a real variable. That Dawkins has not understood the issue in these terms is suggested by his discussion of how 'helpful' he finds 'the backwards way of looking at the past 'experience' of genetic replicators'(1982:95). The 'backwards way' means looking after the fact at what actually happened to tokens of some arbitrary DNA sequence. This

supposedly reveals that the sequence really does have an average fitness value. This idea strongly suggests that Dawkins conceives the problem with tiny replicators like the selfish nucelotide as a mere practical difficulty: it is hard to work out the average fitness value of such a context-sensitive unit. Hence 'while it is not strictly incorrect to say that an adaptation is for the good of the nucleotide, it is not helpful to do so.' (1982:91). But the problem with the selfish nucleotide and other arbitrary DNA sequences is not that their average fitness is hard to calculate, but that it is an artefact. That is why it can only be calculated post-facto ('the backwards way') and why it has no explanatory force.

To really meet the first constraint and avoid the selfish nucleotide *reductio* it is necessary to distinguish two senses in which genes may be 'consistently' on the wining or losing side. George C. Williams defines genes as DNA segments for which there is 'a statistical bias in the relative rates of survival of alternatives.' (Williams 1966):22. This is clearer than Dawkins's slogan about being a consistent loser, but it could still be misread as requiring nothing more than a difference in the *actual* frequency with which some sequence survives. Genes are sequences that can consistently lose in the same sense that an unlucky gambler can consistently lose when tossing a fair coin. Defining fitness in terms of actual success, however, has been widely and justly criticised (Mills and Beatty 1979; Sober 1984a; Sober 1984b). The fact that a gene is 'unfit' in this actualist sense does not explain why it lost, any more than a gambler being a 'loser' explains why they lost a fair coin toss. Obviously, the sensible reading of Williams's definition is a statistical bias in the *expected* frequency with which some sequence survives. There must be some property of the DNA sequence that causes it to win or lose consistently,

just as a weighted coin might consistently fall heads. Only such a nomological fact can genuinely explain the fate of the sequence.

There are just two ways to meet the first constraint. The first is to insist that S-genes be units of molecular function, such as CMG's. The second is to contextualise arbitrary DNA sequences into a setting in which there really are nomological facts about their relative fitnesses. Dawkins's 'backwards way' seems appealing only because it is not clearly distinguished from a real process of contextualisation. Dawkins proposes to take any sequence that actually has survived and to ascribe to it a propensity to survive in its actual historical range of environments. This fails to distinguish between nomological and accidental outcomes, and leads to the amusing consequences described in the last section. A real process of contextualisation ascribes a propensity to survive to an arbitrary DNA sequence in a specific context in which it really is more likely to be reproduced than other sequences. Even the 'selfish nucleotide' theory can be made to work if we use a real process of contextualisation. In the right codon of the right CMG of the right organism in the right environment the single nucleotide T is more likely to be reproduced than an A in the same position (Figure 1).

Insert Figure 1 about here.

The concept of a contextualised gene has been explored recently by Eva Neumann-Held (Neumann-Held 1997). She suggests that the real molecular correlate of a Mendelian gene is not a classical molecular gene, but a DNA sequence plus the whole complex of resources which surround it when it occurs in certain contexts in a lineage of organisms. It is this whole

contextualised system that recurs from one generation to the next and which accounts for the pattern of inheritance at the phenotypic level that the Mendelian gene concept was introduced to explain. Neumann-Held calls this the 'constructionist' gene concept. The constructionist gene concept will not appeal to the hard-line gene selectionist, even though it will allow arbitrary sequences to count as genes. The contextualised gene in Figure 1 is not the bare nucleotide T, but *T in this context*. Contextualised gene selectionism does not predict that T will proliferate over A, but only that T in this context will proliferate over A in this context. This prediction does not fit the Dawkinsian vision of evolution as a struggle between naked replicators. At the core of that vision is the idea that similar DNA sequences have a common interest in the perpetuation of their kind: genes of a feather flock together. This is clear in Dawkins's ultimate rejection of the selfish nucleotide theory on the grounds that: 'It becomes downright misleading if it suggests to the student that adenine at one locus is, in some sense, allied with adenine at other loci, pulling together for an adenine team. … [Nucleotides are] indifferent to the fate of their exact replicas at other loci.' (1982: 91). The contextualised gene concept would imply that all arbitrary DNA sequences are like single nucleotides in this respect. They play for different teams when they transfer between contexts.

The hard-line gene selectionist needs genes to be identifiable independently of context but also to have some real, nomological properties in virtue of which they can be selected. We suggest that the best, and perhaps the only, way of achieving these two goals simultaneously is to demand that S-genes be units of molecular function. In the rest of this paper we will assume that S-genes are classical molecular genes (CMGs). Unlike arbitrarily-chosen

sequences, each CMG has been selected for its ability to produce, in the cellular context of the lineage of organisms in which it is found, at least one molecule that carries out a specific catalytic task. Thus, CMGs have effects that can legitimately be extrapolated to new contexts, at least within the lineage of organisms in which they occur. CMGs will meet the first constraint.

In this section we have shown that the arbitrary sequences apparently licensed by the evolutionary gene concept (EGC) do not meet the first constraint and so cannot be S-genes. S-genes must be units of molecular function. While CMGs may not be the only units of molecular function (isolated regulator or promoter sequences may count too), most of Dawkins's examples of real genes in his many writings are either CMGs or complexes of CMGs. If we can show that CMGs typically do not meet the second constraint — that tokens of a type must be in competition with one another — then we will have shown that genuine S-genes are few and far between.

6. Meeting the Second Constraint — Molecular Taxonomy

The second constraint on any meaningful unit of genes selection is that tokens of a single type of S-gene must be in competition with one another. To fully assess gene selectionism, therefore, we need an account of molecular taxonomy that captures the full range of options for sorting genes into types. Fortunately, taxonomic issues have been extensively worked through in the debate over organismic species, leaving some well-developed concepts for us to apply to genes. Principles for taxonomising biological systems are usually embodied in so-called 'species concepts': definitions of when two organisms are members of the same basal taxon. Most species concepts embody a

taxonomic philosophy that can be applied to related problems. For example, the principles used to sort organisms into evolutionary lineages can also be used to define categories in anatomy (Darwinian homology). Similarly, both organisms and their parts can be classified by pure similarity or by their ecological roles (functions). In this section we explore how the various current approaches to taxonomy can be applied to sort DNA sequences into types. We can then ask whether any of these approaches generates gene types that meet the second constraint, and so may be suitable S-genes around which to develop a theory of gene selection.

There are a plethora of species concepts in the current literature. A recent review by R.L Mayden counts 22, although some of these are very close relatives (Mayden 1997). In another review in the same volume, David L. Hull sorts species concepts into three groups (Hull 1997). Phenetic species concepts aim at a universal, theory free characterisation of species based on some notion of similarity. A second group of species concepts, including Ernst Mayr's biological species concept (BSC), G.G. Simpson's evolutionary species concept (ESC) and A.T. Templeton's cohesion species concept (CSC), define species in terms of the evolutionary forces that create and sustain them. The BSC specifies that the force involved is gene exchange, the ESC leaves the forces largely unspecified and the CSC provides fairly detailed operational guidelines for identifying species produced by a range of possible forces. Finally, phylogenetic species concepts define species in terms of the patterns of ancestry and descent which it is the aim of cladistic or phylogenetic systematics to uncover.

We adopt a very similar classification to Hull's. We divide theoretically committed species concepts into those related to the BSC and those related to Leigh Van Valen's ecological species concept (EcSC) (Van Valen 1976). The EcSC defines species in terms of the ecological niches they occupy and the effect of niches in maintaining the characteristics of species. We separate these two approaches to reflect the differences between the evolutionary forces operating in sexual and asexual taxa. Thus, for example, Templeton's popular cohesion species concept defines a species as 'the most inclusive group of organisms having the potential for genetic and/or demographic exchangeability' (Templeton 1989):25. Asexual species, at least in eukaryotes, cannot achieve cohesion through genetic exchange and so must rely on demographic exchange. These two mechanisms suggest two different ways of defining species of genes. We evaluate one of these as the genic equivalent of BSC, the other as the genic equivalent of EcSC.

Insert Figure 2 about here.

Three of the main taxonomic methodologies — phenetics, phylogenetics, and the ecological species concept — can be applied to genes just as they can to organisms and the phenotypic traits of organisms. These applications are outlined in this section. The application of the biological species concept to molecular genes proves more problematic and is deferred to the next section. In the penultimate section of the paper we return to the ecological species concept and its genetic applications.

Insert figure 3 about here

DNA sequences can be classified phenetically, by sheer similarity of sequence. They can also be classified phylogenetically (cladistically) counting only those features likely to indicate patterns of ancestry and descent amongst gene tokens. Just as in organismic evolution, we can distinguish genuine evolutionary homology from physical similarity which exists for other reasons or by chance (homoplasy or analogy)[5]. Finally, sequences can be classified ecologically, in terms of either their molecular or phenotypic functions. This is clearly seen in the case of null alleles. Several different mutations can cause equivalent functions (or disruptions of function). The allele that produces Myotonic Dystrophy, for example, is any one of a number of sequence of between 50 and 200 trinucleotide repeats (the 'normal allele' being any of several sequences between 5 and 27 repeats) (Brook, McCurrach et al. 1992). As would be expected of an ecological classification, this generates a genic equivalent of convergence or homoplasy. Just as birds and insects share the common feature of having wings, DNA sequences may be 'the same gene' without recent common ancestry, in virtue of their ability or inability to perform the same functional role. Neither the 'normal' nor 'disease' forms of the MD gene form a clade. This form of classification, by catalytic function of gene products, represents the historical rationale of the classical molecular gene concept. There is a continuous line of descent from the 'one gene - one enzyme' hypothesis, through the 'one gene - one polypeptide' concept to the present more complex state of affairs in which gene identity changes as theoretical focus shifts from one stage of post-transcriptional processing to another.

---

[5] The fact that the homology concept is so readily identified with raw physical similarity in the molecular literature is probably a sign that the methodological battles through which systematics achieved its current state of relative conceptual clarity will have to be fought again as molecular systematics matures as a discipline. See: (Hull 1988).

It is clear that classical molecular genes can be, and regularly are, classified into types using all three of the methodologies used to group individual organisms into species. However, this need not group sequences into populations within which there is competition. All three taxonomic methodologies are also used to classify organisms into higher taxa like genus and family, and individuals do not compete merely by being congeneric. It is also unclear that all populations of a species must participate in a single process of competition. The feral wallabies in Yorkshire may not be in competition with their conspecifics in Australia. These considerations make it an open question whether the gene types generated by modern taxonomic principles represent populations within which there is competition and selection. Robert Brandon has expressed the requirements for selection using the notion of *selective environment* (Brandon 1990). Brandon pares down the innumerable factors in the biotic and abiotic surroundings of an organism (*external environment*) in two successive steps: the *ecological environment* is the subset of these factors that affect the reproductive output of individuals in the population, and the *selective environment* is the subset of factors from the ecological environment that affect reproductive output differentially, favouring some variants over others. According to Brandon, it makes sense to interpret change in the proportions of different variants as a process of selection if the variants share a selective environment: a set of factors common to them all but to which they respond differently. Variants will then be selected for their relative ability to cope with that selective environment. In this circumstance, and only in this circumstance, does it make sense to ascribe adaptation to competition between individuals. Thus if gene types are to meet the second constraint, that tokens of a type are engaged in selective

competition, then these types must define populations of gene tokens which face a single selective environment. The mere fact that one variant on a type has more descendants than another is not enough.

The phylogenetic and phenetic classifications of genes actually made by molecular biology typically define groups too far up the taxonomic hierarchy to be subject to a single process of selection. Most phylogenetic groupings (clades) do not face a common selective environment, as Dawkins has noted: 'Lions and antelopes are both part of the class Mammalia, as are we. Should we then not expect lions to refrain from killing antelopes, 'for the good of the mammals'? Surely they would hunt birds and reptiles instead, in order to prevent extinction of the class. But then, what of the need to perpetuate the whole phylum of vertebrates?' (Dawkins 1976):11. In the same way, gene families like the globins do not collude for the replication of 'their own kind'. Each of us probably has an HLA-A allele that is more similar in sequence to an HLA-A allele of some chimpanzees than to those of some of our friends (McAdam, Boyson et al. 1995; Watkins 1995). It is unlikely that all of these HLA-A alleles participate in the same selective process. The situation for phenetic gene species is even clearer. Copies of the same DNA sequence may be spread throughout the genome of an organism and across the genomes of different organisms. We conclude, therefore, that neither the phenetic nor cladistic classifications commonly encountered in molecular biology define gene types that meet the second condition for being an S-gene. It is not typically the case that variants on a single type of phenetic or cladistic gene are in competition with one another. There is no reason to expect competition within these types, because having a physical similarities or having common

ancestor are not properties that connect in any principled way with the property of occupying a particular selective environment.

The project of applying the Ecological Species Concept (EcSC) to define S-genes seem much more promising: the EcSC can be used to define the groups within there is competition for living opportunities. The idea that superior sequences of base-pairs replace their rivals through competition for places in a limited number of future genomes has been central to gene selectionist thinking. If gene types can be defined through such competition, they should be S-genes. But many of the ecological groupings of genes made by molecular biology are actually much too far up the taxonomic hierarchy to define populations within which there is a single selective process. Grouping together all genes that are involved in the production of serine proteases is like grouping together all beaks: while beaks form a single ecological or adaptive category, not all beaked organisms are in competition to have the best beaks. The beaks of squid are not in competition with the beaks of ostriches. It would be possible to measure the relative change in the number of beaks extant in pelagic cephalopods and in savanna-roaming ratites, and to conclude from this that the beaks of the fastest-growing category are selectively superior, but this would be frivolous. However, this does not mean that smaller ecological types might not define populations within which there is competition and selection, as we discuss below.

The other promising way to group gene tokens into populations within which there is genuine selection is the Biological Species Concept (BSC). The BSC seeks to define the groups within which there is competition for a limited number of breeding opportunities, another idea which has been

historically central to gene selectionist thinking. We explore this possibility in the next section, and the idea of ecological competition in the penultimate section.

7. Genes and the Biological Species Concept

According to BSC gene selectionism, S-genes are populations of classical molecular genes whose members actually or potentially recombine at meiosis. These gene types would meet the second constraint because recombination creates a limited set of 'breeding opportunities' for which tokens of the type compete. This constitutes a common selective environment in which variants better able to compete for these opportunities will be selected. Recombination would also keep these gene 'species' relatively homogeneous, just as, according to the BSC, interbreeding maintains similarity within species of organisms. BSC gene selectionism should appeal to hard-line gene selectionists because of the similarity between its definition of an S-gene and the evolutionary gene concept (EGC) favoured by Williams and Dawkins.

There are three major problems with BSC gene selectionism. First, the usefulness of the genic BSC for thinking about evolution is severely limited because the BSC concept of 'same gene' depends on karyotype, and karyotypic change is a major component of evolutionary change. As noted above, loci are well-defined for periods of laboratory time (especially in inbred strains) but can be very hard to identify across evolutionary time because deletions, duplications, translocations, and inversions are common. Hence, even if BSC gene selectionism were workable, it could not explain major evolutionary change because its units are not well-defined across major stretches of

evolutionary time. This objection applies to many versions of the original EGC as well, and it is surprising that it has not been more widely noticed.

Second, the gene types defined by the BSC are not the genes commonly discussed in molecular biology — the globin gene family or the serine proteases — or in evolutionary biology — the gene for sickle-cell anaemia in humans or the gene for melanism in *Biston betularia*. The fact that the 'same gene' in almost all the senses discussed so far can occur at many different loci throughout a genome is sufficient to establish that these gene classifications do not pick out BSC gene types. You can't recombine if you never meet: thus, with the exception of transposons, members of a BSC gene type must share a single locus. It might be objected that the recombination machinery is not picky: any two sequences that are sufficiently similar (in a phenetic sense) may recombine with one another, even if they are at different loci. This illegitimate recombination typically duplicates or deletes the region of chromosome between the loci, leading to unfit or inviable offspring. The BSC assigns parents involved in infertile crosses to different species and will assign sequences that recombine in this way to different BSC gene types.

In fact, under the BSC it is unlikely that genes in different species of organisms can ever count as the same gene type. Your H4 histone genes, for instance, would probably recombine quite happily with those of pea plants (the protein sequences of which differ by only two amino acids), but there are fairly insurmountable reproductive barriers that prevent them from doing so. This is a relatively minor problem, however, since most gene selectionists conceive gene selection as primarily occurring within a single species of organism.

The significance of the mismatch between BSC gene types and the gene types generally discussed in molecular and evolutionary biology is considerable. we can already conclude that the claims of BSC gene selectionists to explain the prevalence of one molecular or phenotypic type over another must be at best indirect. They explain the prevalence of these types only because the tokens of these types are also tokens of other types and the prevalence of those other types can be explained by gene selectionism. The typical gene selectionist story, in which a variant of a gene with a certain molecular or phenotypic function outcompetes other variants in virtue of its superior ability to perform that function is not a story that can be told by the BSC gene selectionist, since that gene type will not be a BSC gene type. An even more serious implication of the mismatch between gene types will become clear at the end of the section.

The third and final objection to BSC gene selectionism is the most significant. The aim of gene selectionism has always been to assimilate the selection of conventional alleles to the selection of selfish DNA, replicating itself without regard for the interests of the organism in which it finds itself. The BSC reconstruction of gene selectionism signally *fails* to accomplish this. Only in unusual cases such as transposons can the selective process of a stretch of DNA be distinct from that of the organism in which it is found. Transposons — 'jumping genes' — produce an enzyme, transposase, that can splice them into other parts of the genome. They can be thought of as 'allelic' to every locus, eliminating normal concerns about what constitutes the 'same locus' over evolutionary time. Because they are 'allelic' to every locus, they can adopt reproductive strategies other than enhancing the fitness of the larger

chromosomal and organismal units of which they are parts. As everyone has always conceded, it makes good sense to think of transposons as engaged in selection process quite distinct from organismic selection. It is, however, more consistent to regard transposons as mini-organisms rather than macro-genes, since the flanking sequences that are required for the transposition and which actually undergo recombination are not part of the transposase coding sequence, and are not themselves translated into protein. Further, transposons typically carry other coding sequences, along with the regulatory apparatus to ensure that they are expressed in particular hosts: the distinction between transposon and virus can be tenuous indeed.

Meiotic drive genes, the paradigm 'selfish genes' of Dawkins's book of the same title, are even more complicated: far more complex than their innocuous name suggests. Meiotic drive occurs when a nuclear DNA sequence is transmitted to offspring at other than 50% frequency (excluding embryonic lethals). The two meiotic drive systems that have been extensively studied are Segregation Distorter on chromosome 2 of Drosophila melanogaster and the t complex on chromosome 17 of Mus domesticus (reviewed by (Hurst, Atlan et al. 1996)). Both transmit at 90-99% strength. These are usually described as genes that subvert the 'fair division' of meiosis. However, neither system actually affects meiosis, and neither system relies on the action of a single gene (CMG). Both apparently act during spermatogenesis by a complex interaction of several diffusible proteins produced by different CMGs that attack a chromosomal target that is not a CMG, but either a non-coding DNA sequence or a chromatin superstructure (Silver pers. comm. as a result of findings in (Ewulonu, Schimenti et al. 1996)). Wild-type sperm have the target and driving sperm lack it, so this process eliminates wild-type

sperm. Segregation Distorter has two main components: Sd (the distorter element) codes for a product that interferes with protein packing, while Rsp (the responder element) is an altered form of the chromosome that doesn't bind the Sd product as the wild-type locus does. The t complex is similar: tcr (the t complex responder locus) is normally deleterious but is rescued by at least four tcd (t complex distorter) CMGs that produce proteins that inactivate wild-type sperm but cannot inactivate tcr sperm (Silver 1985; Silver and Buck 1993). In both cases, the necessary genes (no one gene alone has the distortion effect) are linked together by one or more inversions that prevent recombination. Both inversions have accumulated homozygous lethal mutants, which is presumably adaptive, since homozygotes are sterile. Meiotic drive, when coupled with deleterious mutations, can lead to local population extinction or can remain at a low equilibrium frequency, apparently increasing the frequency of the meiotic drive 'gene' (actually, about a quarter of a chromosome in the case of the t complex) at the expense of the fitness of the organisms. In either case, this is a selection process between types of sperm, not between types of genes. t sperm outcompete + sperm in every situation in which they occur together (i.e. heterozygous males); a side-effect of this selection process is that developmental systems fostering suitable environments for this selection process to occur (i.e: heterozygous males) become more common. There is no need to explain the population-level effects as the triumph of selfish genes over cooperative genes.

The tendency of gene selectionists to ignore the real structure of transposons and of so-called meiotic drive genes helps foster the false impression that every gene is 'potentially' a subversive, selfish gene. In fact, subversive behaviour is a major evolutionary achievement. Even the strongest

opponents of hard-line gene selectionism have always recognised that intra-genomic selection is real and important (Lewontin and Dunn 1960; Lewontin 1970), but it is not merely a striking case of the same process that explains phenotypic evolution, as the gene-selectionist would have us believe. In marked contrast to transposons, there is only one way for a normal allele to increase its recombination opportunities: by increasing the fitness of the entire organism. Gene selection of this sort cannot act in opposition to the selection of the phenotypic traits produced by molecular genes.

It is at this point that the fact that BSC gene types are not the gene types of molecular or evolutionary biology assumes its true significance. A selective explanation of the prevalence of normal alleles cites their contribution to the larger system of which they are part. Two methods of gene-classification seem relevant to such an explanation: phylogenetic, in terms of descent, and ecological: in terms of catalytic activity. What BSC gene selectionism seems to offer is a way of classifying genes that systematically misses the gene-types that are the focus of the relevant selection process. At normally segregating loci, BSC genes compete in virtue of the catalytic activity of their products. But tokens of BSC gene types have these effects because they are, coincidentally, tokens of gene types which cut across the boundaries of the BSC types. Hence, although the gene tokens recognised by the BSC have causal powers, the BSC gene types are not the types mentioned in the causal generalisations that explain the selection of these gene tokens. The selection of BSC gene types is a side-effect of the selection of the functional gene types with which they coincidentally overlap. Hence BSC gene types are not the right types with which to explain the selection of the gene tokens. The fact that the genes which proliferate are genes of a given BSC type is an *accidental* consequence

of the selection processes that are operating. With this in mind we turn to the more promising alternative of EcSC gene selectionism.

8. Genes and the Ecological Species Concept

Ecological classifications are often too large to define a population facing a common selective environment. They can do so, however, when individuals compete for a limited number of 'living spaces' in a habitat. The ecological species concept (EcSC) defines a species as a group of organisms whose members share an adaptive niche and who can replace one another's descendants if they find more efficient ways to occupy the niche. A gene selectionist might argue that each organism corresponds to just such a 'living space' and that CMGs compete for representation in organisms. After all, paradigmatic selection processes, such as selection for antibiotic-resistance genes, can take place in asexual populations of bacteria where the BSC can have no application. Nevertheless, variant genes that can perform a particular catalytic task more efficiently will replace those that fulfil it less efficiently.

Like the BSC, the EcSC offers an explanation for the cohesion of genic 'species'. Genes such as histones remain the same over geological time not because they recombine with one another but because they occupy a catalytic niche: any loss of function reduces their reproductive prospects.

Another appeal of the EcSC approach to classifying genes is that it has the same basic rationale as the classical molecular gene concept: one gene - one gene product. We have briefly mentioned some of the problems that face this idea. The DNA structures usually identified as CMGs produce transcripts

which are extensively modified in a way that depends on their genomic context and on cellular conditions. These facts makes it unclear how much context to include in a CMG token. They also imply that token CMGs with the identical primary transcripts may have quite different final products. So the identity conditions for CMG types will change if we focus on different stages of post-transcriptional processing. One way to confront these difficulties is to say that two CMGs are the same gene if they have the same norm of reaction across a range of genomic and cellular contexts. This definition attempts to capture what it is for two sequences to be exactly the same type of CMG. Certain differences — in the third position of codons, for example — might be tolerated because they do not cause differences in protein products, but CMGs of the 'same type' by this test are ecologically indistinguishable.

Both molecular biology and gene selectionism will require much broader, but similarly defined, notions of 'same gene'. The simplest example of a looser ecological notion of gene identity in molecular biology occurs with null alleles. The 'gene for albinism' in a species may be any one of a number of dysfunctional variants on any one of a number of CMGs involved in producing skin pigmentation. Gene selectionism will need a looser notion of 'same gene' since the whole idea is that tokens of a gene type may compete to fulfil the same catalytic niche in a limited number of organisms. The gene tokens cannot be identical, since they compete by performing the catalytic task more or less well. Like individuals of an asexual moss species competing for a limited number of ecologically suitable growing sites, they must be similar enough to fill the same niche but varied enough to differ in how well they fill it.

The ecological notion of 'same gene' that is probably most relevant to gene selectionism is the notion of developmental interchangeability. If we consider a living opportunity for a CMG to be a location on a chromosome of a viable organism, then any two CMGs that could be substituted for one another in a viable organism can be considered ecologically equivalent, in the sense that they are able to occupy the same 'living site'. If substitution of one CMG for another would make the organism inviable, then the two CMGs cannot be of the same EcSC type. Some elaboration of this definition will be required because some genes have overlapping or redundant functions. The myogenic regulatory factor MyoD is involved in muscle development in mice from mid-gestation onwards. However, knockout mice in which this gene is disabled develop more or less normally, because in the absence of MyoD transcription another myogenic regulatory factor, Myf5, which is usually expressed only in early embryogenesis, continues to be transcribed (Weintraub 1993). This makes the 'inviability' test useful only if it produces a positive result. A more adequate definition of an EcSC gene type might be that two CMGs are of the same type if substituting one for another at a site does not eliminate the catalytic function of the site. We propose this as a definition of S-gene type for EcSC gene selectionism.

The question now is whether these types meet the second constraint. Is it the case that tokens of an EcSC gene type are in (ecological) competition to occupy a limited number of living sites offered by organisms in the next generation? Is it the case that variants on an EcSC gene type that perform their catalytic task more efficiently will be selected over the 'wild type'? The answer to both questions is 'not in general', as can be seen from an example. Different opsins with absorption maxima at 570 nm and 530 nm allow humans (and other

primates) to see red and green respectively. In monkeys, the gene for the red opsin and for the green opsin are allelic to one another and X-linked, so males can see red or green but not both. In apes (including humans), a gene duplication has occurred such that males have both red and green opsins, and can see both red and green (Deeb, Jorgensen et al. 1994; Shyue, Li et al. 1994; Jacobs 1996). There is no reason, other than historical accident, for the red and green opsins to compete with one another in primates. This is true quite generally: competition between tokens of EcSC gene types occurs only due to historical accidents of chromosomal placement. Tokens of a single type need not occupy the same locus: if types are defined by functional interchangeability then tokens may occur anywhere in the genome. When two CMGs have related but slightly different functions, an organism that expresses both functions is at a competitive advantage. EcSC gene tokens are not competitors, because the reproductive prospects of both variant and wild type are optimised by the reproduction of both types. Organisms that have duplicated genes may have all the functionality of what were once competitors. The same is true for two wild-type tokens, because of the contribution of gene duplication to the developmental canalisation in the face of environmental and genetic variation.

The gene selectionist might try to salvage ecological competition by restricting S-genes defined by ecological role to a single locus. Two CMGs are the same S-gene if they are the same EcSC gene type *and* the same BSC gene type. There are at least two objections to this proposal. The first of our three objections to BSC gene selectionism applies with full force. S-genes of this kind are simply not well-defined over any significant portion of evolutionary time because of karyotypic change. The third objection to BSC selectionism applies in a

modified form. EcSC gene types are selected for their contributions to metabolic efficiency. This selection process is indifferent to locus. In sexual organisms it is clearly a single selection process that favours possession of a particular functional gene irrespective of locus. If a yeast 23S ribosomal RNA gene mutates so as to confer resistance to some antibiotic, it does not matter which of the hundreds of copies per genome mutated. The same (convergently altered) form is doing the same job even if it is on chromosome 1 in one individual and chromosome 2 in another. There is also clearly selection for EcSC gene types (sometimes even the *same* types) in asexuals, where the BSC cannot be applied. So the composite EcSC-BSC gene-types face the third objection to BSC gene selectionism. They cannot be the right way to understand the selection process in question since they exist only in some instances of that selection process. The composite gene selectionist proposal mistakes the particular for the general, the accidental for the essential.

This final, EcSC, version of gene selectionism fails dismally. It begins with the intuition that since organisms compete by functioning more or less efficiently, the genes that underlie this functionality must be in competition with one another. This is simply a mistake. The genes that underlie functionality do better by complementing than by competing with one another.

9. Why Does Gene Selectionism *Seem* Coherent?

In this paper we have tried to rigorously formulate the idea that there is selection *for* rather than merely selection *of* (Sober 1984a; Sober 1984b) individual genes. We have tried to show that on no definition of 'gene' is this

generally true. The only genes that meet both constraints on S-genes are the classic 'selfish DNA' cases — transposons and meiotic drive complexes. Our argument relies only on uncontroversial facts of molecular biology, the standard principles of biological taxonomy and one of the most widely accepted accounts of what constitutes selection and competition. This raises the obvious question of why gene selectionism has such appeal. We suggest that gene selectionism is in reality atomistic *phenotypic* selectionism disguised by a specious appeal to genetics. Atomism about phenotypes, the assumption that traits are sufficiently independent of one another for selection to operate on traits rather than trait-complexes, has an obvious appeal as part of the adaptationist program in biology. The appeal to genetics functions to illicitly brush under the carpet some objections to atomism.

The connection between gene selectionism and atomism is suggested by the use of the 'extended phenotype' cases as arguments for gene selectionism. In extended phenotype cases, traits of one organism are selected for the benefit of another. The fungus Entomophthora muscae infects and kills domestic flies. It causes dead females to develop features such as a distended abdomen, which are sexually attractive to male flies, so that the necrophile males become infected with the fungus (Moeller 1993) Examples like this are meant to demonstrate that a trait (distended abdomen) can be selected without selection acting for the benefit of the phenotype that displays the trait (the fly). Dawkins argues for gene selection on the grounds that it makes sense of such cases and, further, shows that the apparent unity of more ordinary phenotypes is an illusion. In reality, the 'genes for' each trait are being selected independently (Dawkins 1982). We are not concerned here to question the heuristic value of the atomistic approach to phenotypes. We merely show

that it receives no support from the idea that it is the 'genes for' the trait that are being selected. Phenotypic atomism must learn to stand on its own feet.

Most gene selectionist explanations concern the 'genes for' some phenotypic trait: genes for camouflaged plumage, for aggressive behaviour or for the extended traits of parasites. These explanations consist of a selective explanation for a phenotypic trait and the claim that it is really the 'genes for' this trait that are being selected. The distended abdomen in the house-fly is selectively advantageous to the parasitic fungus, but it is the 'genes for' the abdomen (which are in the fungus) that are really being selected. Explanations of this kind have a hidden agenda and a flawed assumption. The hidden agenda is to sweep under the carpet the problems of isolating one trait from the others with which it is developmentally and functionally entangled. This is done by supposing that the 'genes for' each trait have an average effect that can be made independent of context and hence of the trait complexes of which each trait is part. The flawed assumption is that the 'gene for' locution is tracking a constant, underlying difference maker, which can be assigned a single, meaningful fitness value, hence avoiding the problem of trivial, artefactual average fitnesses described in section four.

The 'gene for' locution has been carefully analysed by Kim Sterelny and Phillip Kitcher (Sterelny and Kitcher 1988). A DNA sequence is 'for' some phenotypic trait if it accounts for variance in that trait in normal genomic and environmental contexts. A gene is a context-sensitive difference maker: it does not make the trait, but all else being equal it makes the difference between one trait and another. Selection acts on these differences, and hence on the gene. Sterelny and Kitcher thus hope to avoid genetic determinism in

developmental biology while allowing gene selectionism in evolutionary biology. The assumption that the 'gene for' locution can track a constant, difference maker during a selection process is flawed because the 'genes for' a phenotypic trait will typically change during selection for or against that trait. This is because causal conditions that become widespread due to selection will tend to disappear from an analysis of variance, whilst those that had no effect in the absence of these conditions will suddenly become prominent. This is nicely demonstrated in a recent paper by H. Frederick Nijhout and Susan Paulsen (Nijhout and Paulsen 1997). They modelled selection for wing patterns in butterflies using a model epigenetic system based on Nijhout's extensive work on the development of wing patterns (Nijhout 1990; Nijhout 1991). Nijhout and Paulsen's model contains six developmental parameters controlling the distribution of a diffusible morphogen. The model assumes that each parameter is determined by a single locus, and that two alleles exist in the population at each locus, one associated with a high parameter value, the other with a low value. Despite the simplicity of this genetic system, they conclude that '… whether a particular gene is perceived to be a major gene, a minor gene or even a neutral gene depends entirely on the genetic background in which it occurs, and this apparent attribute of a gene can change rapidly in the course of selection on the phenotype.' (1997: 401-3, See figure 4).

Insert figure 4 about here.

Gene selection seems plausible because it uses the 'gene for' locution to suggest a constant difference maker underlying the variance in the phenotype and acting as the beneficiary of adaptation based on that variance. In fact, the

gene for locution will pick out in turn different DNA sequence that account for the major part of the variance in a trait as selection proceeds. This fact goes unnoticed because the 'gene for' locution is expounded using atypical, pathological cases in which the complexities of the gene-phenotype relation can be ignored. Everyone is familiar with 'genes for' disease phenotypes. These involve a major defect in an evolved gene whose normal role involves connections to many other genes, either by direct protein/protein interaction or downstream in vital pathways. Such disease genes include genes for albinism, melanism and other pigment changes caused by defect in a gene that makes pigment, suppresses pigment formation or breaks pigment down; dwarfism caused by enzyme defect; and the most famous case of all: phenylketonuria and its variants caused by loss of an enzyme. These pathological genes impair normal development to such an extent that they dominate variance in the phenotypic traits they affect under almost any background conditions.

We conclude that the form of gene selectionism that has gripped the popular imagination is really atomistic phenotypic selectionism. In practice the 'genes for' under discussion are usually fictions and the purpose of these fictions is to sidestep concerns about treating phenotypic traits atomistically. This goes a long way to explain the popularity of gene selectionism. Phenotypic atomism is the 'dream ticket' of the adaptationist program in biology. We are not concerned here to question the utility of phenotypic atomism, only to point out that the support it receives from 'gene selection' is specious. Note also that a 'gene-for selectionism' which was actually concerned with genes would face the problems of EcSC gene selectionism discussed in section eight, since these gene types would be functionally defined.

10. Conclusion

Gene selectionism enjoys its present popularity because of its simplicity. Unfortunately, this is not the simplicity of elegance, but the simplicity of crudeness. What could be simpler than assuming every instance of 'gene' to refer to the same entities? Genes produce proteins; genes store information; genes cause disease; genes have functions; genes compete with one another. Again, the reality is more complex. As has been noted many times in the last thirty years, different people mean different things by 'gene'(Kitcher 1982; Falk 1984). The genes that are selfish lack functions: the genes that have functions lack selfishness. No one gene can carry the weight of the semantic associations with which the word 'gene' is laden.

What could be simpler than averaging away all the vagaries of chance to arrive at the true selection coefficient for any sequence? By eschewing the nomological in favour of the accidental, genic fitness can apparently be reduced to a simple matter of history. But at the same time, it is robbed of its explanatory force. The junk DNA carried by dodos had no effect on their demise: they were not 'bad genes'. They were merely unlucky, despite their low 'fitness values'.

We have found a number of defensible doctrines that could be described as gene selectionism. The contextualised gene concept can make sense of the idea that arbitrary sequences — even single nucleotides — are selected in specific contexts, but it does not licence them to carry their fitness values away from that context or to average it across contexts. The idea that the alleles at a locus compete to occupy loci in the next generation is true in a sense, but not in one that makes the selection of normally segregating alleles fundamentally

the same as the selection of 'selfish DNA'. Finally, of course, there are real cases of intra-genomic selection, such as transposons, but these do not account for phenotypic evolution. None of these doctrines fulfils gene selectionism's promise of a single, simple unit of selection underlying all evolutionary processes. The units of selection debate needs to shift its focus to processes, not parts; to systems, not structures. Only then will attention be drawn to the units, rather than the eunuchs, of selection.

R. D. Knight,
Ecology and Evolutionary Biology,
Princeton University,
Princeton,
New Jersey 08544
rdknight@princeton.edu

P. E. Griffiths,
Philosophy,
Otago University,
P.O. Box 56, Dunedin,
New Zealand
pgriffiths@rivendell.otago.ac.nz

Figures

Organism	Normal	Carrier	Anaemic
reduces to			
Genotype	HbA/HbA	HbA/HbS	HbS/HbS
reduces to			
Amino Acid at Position 6 of β-Globin Peptide	Glu/Glu	Glu/Val	Val/Val
reduces to			
Codon at Position AUG+6 of β-Globin mRNA	GAR/GAR	GAR/GUR	GUR/GUR*
reduces to			
Base at Position ATG+17** of β-Globin Cistron (antisense strand)	A/A	A/T	T/T
Fitness†	0.88	1.00	0.14

Figure 1: The Selfish Nucleotide Theory in Action.

*R is the symbol for either of the purines, A or G.
**Note that this is not an absolute chromosomal position, but is relative to an external promoter sequence.
†From (Bodmer and Cavalli-Sforza 1976), cited in (Ridley 1993).

(Generalised) Phenetic Species Concept
Species are clusters of organisms that have similar phenotypes.

(Generalised) Phylogenetic Species Concept
Species are separate lineages of organisms — groups within which there is a single pattern of ancestry and descent and between which there are discontinuities in the pattern ancestry and deescent

(Generalised) Ecological Species Concept:
Species are groups of organisms that share the same ecological niche.

(Generalised) Biological Species Concept:
Species are groups of organisms that can interbreed with one another.

Figure 2. Four Approaches to Taxonomy

Phenetic Species Concept:
Gene types are sets of gene tokens that have similar sequences.

Cladistic Species Concept:
Gene types are monophyletic lineages (clades) of gene tokens.

Ecological Species Concept:
Gene types are sets of gene tokens that perform similar functions.

Biological Species Concept:
Gene types are sets of gene tokens that can combine with one another during meiosis.

Figure 3. Applying Species Concepts to Genes

Scan of Fig 3. from (Nijhout and Paulsen 1997) here.

Figure 4. **A.** Response of the phenotype to truncating selection for a larger value. In each generation individuals with phenotypic values below the mean minus one-half of the standard deviation of the previous generation were removed from the breeding population. **B.** Genetic response to truncating selection on the phenotype. **C.** Genetic correlations of the six developmental parameters with the phenotype during selection. Note that genes at the*Decay, Threshold* and *Background* loci are not 'genes for' the trait early in the selection process but are major genes later in the process (From (Nijhout and Paulsen 1997))

References

Bodmer, W.F and Cavalli-Sforza, L.L [1976]: *Genetics, Evolution and Man*, San Francisco, W.H Freeman.

Brandon, R [1982]: 'The levels of selection', in P. Asquith and T. Nickles (eds), *Proceedings of the Philosophy of Science Association*, Vol. 1, pp. 315-322.

Brandon, R [1990]: *Adaptation and Environment*, Princeton, Princeton University Press.

Brook, J.D, McCurrach, M.E, et al. [1992]: Molecular basis of Myotonic Dystrophy: Expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member, *Cell*, **68**, pp. 799-808.

Dawkins, R [1976]: *The Selfish Gene*, Oxford, Oxford University Press.

Dawkins, R [1982]: *The Extended Phenotype*, Freeman.

Dawkins, R [1995]: Reply to Lucy Sullivan, *Philosophical Transactions of the Royal Society of London*, **B 349**, pp. 219-224.

Deeb, S.S, Jorgensen, A.L, et al. [1994]: Sequence divergence of the red and green visual pigments in great apes and humans, *Proceedings of the National Academy of Sciences*, **91**, 15, pp. 7262-6.

Ewulonu, U. K, Schimenti, J, et al. [1996]: Targeted mutagenesis of a candidate t complex responder in mouse t haplotypes does not eliminate transmission ratio distortion., *Genetics*, **144**, pp. 785-792.

Falk, R [1984]: The gene in search of an identity, *Human Genetics*, **68**, pp. 195-204.

Griesemer, James [Forthcoming] *Reproducers, Replicators and the Units of Evolutionary Transition*, pp**.** 2-21.

Griesemer, J. R [1992]: 'The Informational gene and the substantial body: on the generalisation of evolutionary theory by abstraction', in N. Cartwright and M. Jones (eds), *Varieties of Idealisation*, Amsterdam, Editions Rodopi.

Griffiths, P.E and Gray, R.D [1994a]: Developmental Systems and Evolutionary Explanation, *Journal of Philosophy*, **XCI (6)**, pp. 277-304.

Griffiths, P.E and Gray, R.D [1994b]: Replicators and vehicles? Or developmental systems?, *Behavioral and Brain Sciences*, **17 (4)**, pp. 623-624.

Hull, D [1988]: *Science as a Process*, Chicago, Univ. of Chicago Press.

Hull, D [1997]: 'The ideal species concept - and why we can't get it', in M. F. Claridge, H. A. Dawah and M. R. Wilson (eds), *Species: The Units of Biodiversity*, New York, Chapman Hall, pp. 358-379.

Hurst, L.D, Atlan, A, et al. [1996]: Genetic Conflicts, *Quarterly Review of Biology*, **71**, 3, pp. 317-364.

Jacobs, G.H [1996]: Primate photopigments and primate color vision, *Proceedings of the National Academy of Sciences*, **93**, 2, pp. 577-581.

Kettlewell, H.B.D [1973]: *The Evolution of Melanism*, New York, Oxford University Press.

Kitcher, P [1982]: Genes, *British Journal of Philosophy of Science*, **33**, pp. 337-359.

Lewontin, R [1970]: The units of selection, *Annual Review of Ecology & Systematics*, **1**, pp. 1-14.

Lewontin, R and Dunn, L. C [1960]: Evolutionary dynamics of a polymorphism in the house mouse, *Genetics*, **45**, pp. 705-722.
Mayden, R.L [1997]: 'A hierachy of species: the denouement in the sage of the species problem', in M. F. Claridge, H. A. Dawah and M. R. Wilson (eds), *Species: The Units of Biodiversity*, New York, Chapman & Hall, pp. 382-423.
Mayr, E [1963]: *Animal Species and Evolution*, Cambridge, M.A, Harvard University Press.
McAdam, S.N, Boyson, J.E, et al. [1995]: Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles, *Journal of Immunology*, **154**, 12, pp. 6421-9.
Mills, S and Beatty, J [1979]: The propensity interpretation of fitness, *Philosophy of Science*, **46**, pp. 263-286.
Moeller, A.P [1993]: A fungus infecting domestic flies manipulates sexual behaviour of its host, *Behavioral Ecology and Sociobiology*, **33**, 6, pp. 403-7.
Neumann-Held, E.M [1997]: 'The Gene is Dead - Long Live the Gene: Conceptualising the gene the Constructionist Way', in P. Koslowski (ed), *Developmental Systems, Competition and Cooperation in Sociobiology and Economics*, Berlin, Springer-Verlag.
Nijhout, H.F [1990]: A comprehensive model for colour pattern formation in butterflies, *Proceedings of the Royal Society of London B, Biological Sciences*, **239**, pp. 81-113.
Nijhout, H.F [1991]: *The Development and Evolution of Butterfly Wing Patterns*, Washington, D.C, Smithsonian Institution Press.
Nijhout, H.F and Paulsen, S.M [1997]: Developmental models and polygenic characters, *American Naturalist*, **149**, 2, pp. 394-405.
Ridley, M [1993]: *Evolution*, Cambridge, Mass., Blackwell Science, Inc.
Sarkar, S [1996]: 'Biological information: A sceptical look at some central dogmas of molecular biology', in S. Sarkar (ed), *The Philosophy and History of Molecular Biology: New Perspectives*, Vol. 183, Dordrecht, Kluwer Academic Publishers, pp. 187-232.
Shyue, S.K, Li, L, et al. [1994]: Intronic gene conversion in the evolution of human X-linked color vision genes, *Moecular Biology and Evoution*, **11**, 3, pp. 548-51.
Silver, L. M [1985]: Mouse t haplotypes, *Annual Review of Genetics*, **19**, pp. 179-208.
Silver, L. M and Buck, C [1993]: The mouse t complex distorter-3 (Tcd-3) locus and transmission ratio distortion, *Genet. Res. Comb*, **63**, pp. 133-137.
Sober, E [1984a]: *The Nature of Selection*, M.I.T. Press.
Sober, E [1984b]: 'Force & disposition in evolutionary theory', in C. Hookway (ed), *Minds, Machines & Evolution, Hookway, C (Ed) CUP*, Cambridge, Cambridge University Press, pp. 43-62.
Sober, E and Lewontin, R. C [1982]: Artifact, cause & genic selection, *Philosophy of Science*, **49**, pp. 157-180.
Sterelny, K and Kitcher, P [1988]: The return of the gene, *Journal of Philosophy*, **85 (7)**, pp. 339-361.

Templeton, A.T [1989]: 'The meaning of species and speciation: a genetic perspective', in D. Otte and J. A. Endler (eds), *Speciation and its Consequences*, Sunderland, M.A, Sinauer Associates, pp. 3-27.
Van Valen, L [1976]: Ecological Species, multispecies and oaks, *Taxon*, **25**, pp. 233-9.
Waters, C.K [1990]: 'Why the Antireductionist Consensus Won't Survive the Case of Classical Mendelian Genetics', *Philosophy of Science Association Proceedings*.
Waters, K [1994]: Genes made molecular, *Philosophy of Science*, **61**, pp. 163-185.
Watkins, D.I [1995]: The evolution of major histocompatibility class I genes in primates, *Critical Reviews in Immunology*, **15**, 1, pp. 1-29.
Weintraub, H [1993]: The MyoD family and mycogenesis: redundancy, networks and thresholds., *Cell*, **75**, pp. 1241-1244.
Williams, G. C [1966]: *Adaptation & Natural Selection*, Princeton, Princeton University Press.
Williams, G.C [1992]: *Natural Selection: Domains, Levels and Challenges*, New York, Oxford University Press.
Wimsatt, W.C [1980]: 'Reductionistic Research Strategies and Their Biases in the Units of Selection Controversy', in T. Nickles (ed), *Scientific Discovery: Case Studies*, D. Reidel Publishing Company, pp. 213-259.