ELSEVIER Stud. Hist. Phil. Biol. & Biomed. Sci. 35 (2004) 647–673

Studies in History and Philosophy of Biological and Biomedical Sciences

www.elsevier.com/locate/shpsc

How biologists conceptualize genes: an empirical study

Karola Stotz [a], Paul E. Griffiths [b], Rob Knight [c]

[a] *Department of History and Philosophy of Science, 1017 Cathedral of Learning, Pittsburgh, PA 15260, USA*

[b] *Biohumanities Project, HPRC, University of Queensland, St Lucia, QLD 4072, Australia*

[c] *MCD Biology, Campus Box 347, University of Colorado, Boulder, CO 80309, USA*

Received 17 September 2003; received in revised form 11 January 2004

Abstract

Philosophers and historians of biology have argued that genes are conceptualized differently in different fields of biology and that these differences influence both the conduct of research and the interpretation of research by audiences outside the field in which the research was conducted. In this paper we report the results of a questionnaire study of how genes are conceptualized by biological scientists at the University of Sydney, Australia. The results provide tentative support for some hypotheses about conceptual differences between different fields of biological research.

Keywords: Conceptual change; Epistemic object; Gene concept; Gene-D; Gene-for; Gene-P

1. Introduction

The philosophy of biology is concerned with those biological debates in which conceptual and empirical issues are so entangled that progress demands both scientific knowledge and the tools of philosophical analysis (Sterelny & Griffiths, 1999, pp. 5–7). The contested and multi-faceted concept of the gene is at the heart of many of these debates. The study of the gene concept, however, poses a challenge

E-mail addresses: kstotz@pitt.edu (K. Stotz); pauleg@pitt.edu (P. E. Griffiths); rob@spot.colorado.edu (R. Knight).

1369-8486/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsc.2004.09.005

to the traditional philosophical practice of conceptual analysis. Traditionally, philosophers have relied on their individual linguistic competence with the corresponding words. When analyzing a concept, the philosopher treats him or herself as a sociolinguistic 'sample of one'. This approach can be extended to technical concepts, since any philosopher of biology should be in a position to consult his or her intuitions as a *biologically literate* sample of one (Neander, 1991). But much of the philosophical literature emphasizes the diversity of conceptualizations of the gene, either over time or between different fields. A sample of one is manifestly not going to reveal the ways in which ideas about the gene *differ* on the basis of differences in theory, training, experience or research focus.

The philosophical perspective that informs the design of this study is one in which scientific categories are conceived as ongoing—and possibly ramifying—projects of deriving empirical generalizations of increasing scope and reliability by adjusting both the extension of those categories, so as to encompass sets of instances with as much in common as possible, and the intension of those categories, so that statements involving the category change their modal status in a way that reflects the centrality of those statements to current theory (Griffiths, 1999). In addition, conceptual change is driven by pragmatic and normative projects that employ the same categories (Griffiths, 2004, Hacking, 1991). This philosophical perspective is congruent with the so-called 'theory view of concepts' in contemporary psychology, according to which a concept is a node in a network of beliefs about the cognitive domain in which its object lies (Medin, 1989). The perspective also resonates with Hans-Jörg Rheinberger's discussion of the gene as an 'epistemic object' in molecular biology—an entity whose name is introduced as a target of research rather than to designate something with which researchers are acquainted (Rheinberger, 1997, 2000)[1] A scientist's grasp of an epistemic object is constituted by the set of experimental practices through which they seek to establish facts about it. Hence differences in the experimental practices used by particular scientists will be reflected in differences in how they apply the concept and in their modal intuitions about the epistemic object.

From this perspective, the conceptual analyst has no alternative but to examine what different biologists say and do. There are a number of ways to achieve this. One is via the history of science, an approach which is extremely valuable and which forms part of the background to this study. Another is by comparing published work from several different scientific fields. Authors in contemporary philosophy of genetics do this, as well as talking to biologists with whom they collaborate. Current debate in the philosophy of genetics is thus biologically well informed and a fertile source of hypotheses and suggestive arguments about the gene concept, its varieties and their uses. However, no individual can be equally well acquainted with the whole spectrum of contemporary biological research. It is

[1] In a recent essay, one of us has argued, *contra* Rheinberger himself, that Rheinberger's insights are consistent with the view that scientific concepts stand in a referential relationship to states of affairs outside the discourses in which they are employed (Griffiths, 2002).

also plausible that the biologists who choose to collaborate with philosophers or to participate in philosophical debates are unrepresentative of the biological community as a whole. Our aim in this study was to approach the topic in a more systematic fashion. We hoped to avoid the inevitable biases that come from having worked in one particular biological field before becoming a philosopher, from collaborating with some particular research group or simply from having a particular interest in one or more fields of research. Although this can only be regarded as a preliminary study, it suggests that a more systematic and quantitative approach to this and other conceptual issues in the philosophy of biology is both feasible and rewarding.

2. Recent work on the concept of the gene

The founders of classical Mendelian genetics were divided as to whether the gene was primarily a postulated explanatory entity or an instrumental device by which to express regularities in the transmission of phenotypic characters (Falk, 1986). Raphael Falk argues that this led to a productive dialectic in which discoveries about the chemical nature of the gene alternated with new functional definitions of the gene at progressively deeper levels of chemical analysis. The initial Mendelian postulate of a close correspondence between gene and trait was replaced by a postulated one-to-one correspondence between gene and enzyme. This in turn was replaced by a one-to-one correspondence between genes and elements of proteins. The 'classical molecular gene concept',[2] which emerged in the 1950s and retains considerable currency today, identifies a gene with a stretch of DNA that codes for one of the polypeptide chains that goes to make up a functional protein. This dialectical development of the gene concept can be interpreted as reflecting a desire to keep the structural and functional definitions of the gene focused on a single entity. When the best structural definition turns out to create units with indeterminate function, structure and function can be brought back into step by using a more proximal description of function: rather than a gene having an indeterminate effect on the phenotype, it has a determinate effect on one of the structural elements that contributes to the phenotype (Griffiths & Neumann-Held, 1999).

The classical molecular gene concept, however, was only a temporary resting point. Despite the fact that the actual molecular underpinnings of the Mendelian gene had only just begun to be uncovered, the classical molecular gene was introduced as a definite structure that has the function, rather than as a functionally defined entity that might or might not correspond to a unique type of structure at the molecular level.[3] The functional conception of the gene, defined by and

[2] References to different 'gene concepts' (for example, evolutionary, Mendelian, classical molecular) have the same problematic status as the ubiquitous references to different 'species concepts' in biology. As far as possible we will write, instead, of various different 'conceptions' of the gene, to avoid vexed issues about 'counting concepts' along axes of conceptual difference across time or across fields.

[3] We are indebted to Raphael Falk for clarification of this part of the discussion.

embedded in the research practices of Mendelian genetics, and the classical molecular vision of the structure of a gene have proved difficult to keep in strict concordance with one another. An initial difficulty with the classical molecular conception is that the actual activity of the gene, and hence its contribution to the phenotype, depends on elements outside the transcription unit. This has led to definitions of the gene, which include the promoter and regulatory sequences that affect whether the gene will be transcribed. In a case the famous lac operon in *E. coli*, these sites are immediately upstream of the site at which transcription is initiated and it is easy to regard them as parts of the gene. In eukaryotes, however, regulatory regions can be distant from the rest of the gene and can be involved in the regulation of more than one gene. It is perhaps unproblematic to regard regulatory regions that are not transcribed into RNA as neither genes themselves nor parts of any other specific gene, although this is certainly a departure from the classical Mendelian conception of a gene as a segment of a chromosome, different allelic forms of which can be tracked via their differing effects on the phenotype. It is less easy to treat actual coding sequences in this way. The one-to-one correspondence between stretches of coding DNA and genes is challenged by the existence of overlapping genes, which share some of the same sequence. Here we see the very same DNA treated as (part of) two different genes because those genes produce different gene products. But there must be additional reasons why these cases are treated in this way, since the ubiquitous existence of introns in eukaryote genes allows several gene products to be made from a single gene by cutting and splicing the primary mRNA transcript in alternative ways.

One response to introns is to use the above-mentioned strategy of retaining a unitary function for each gene by moving the function closer to the DNA itself. If a gene is defined as the stretch of DNA coding for a single primary mRNA transcript, rather than a single polypeptide, then a gene containing introns can still be defined by a single gene product. Another alternative is to abstract away from the details of the various spliced transcripts to obtain a single feature to associate with the gene from which they are all transcribed. For example, it has been argued that the whole family of transcripts preserves the linear order of codons, omitting different ones but never reversing the order or inserting additional codons (Epp, 1997). However, the phenomenon of mRNA editing, in which individual bases that do not correspond to bases in the DNA are inserted into the mRNA transcript, means that not all 'gene products' have even this abstract relationship to the DNA from which they originate. Another class of problems for the classical molecular gene conception arises because of transplicing, the phenomenon in which mRNA transcripts from several different loci are brought together and spliced into a single mRNA before being translated into a single 'gene product'. Cases involving transplicing can be treated as a single gene split between several loci, as a process for deriving a single product from more than one genes, or, where one transpliced element is somehow subordinate to the other, as an instance of a single gene with a distant, transcribed regulatory region (Fogle, 2001).

One response to the variety of structural and functional units that can be usefully defined in contemporary molecular genetics is pluralism. Falk writes: 'Today

the gene is not *the* material unit or *the* instrumental unit of inheritance, but rather *a* unit, *a* segment that corresponds to *a* unit-function as defined by the individual experimentalist's needs' (Falk, 1986, p. 169). Such pluralism is not necessarily a criticism of the current state of affairs. Falk is unsure if the current ambiguities of the gene concept will prove as helpful as the ambiguities that proved so productive in classical Mendelian genetics. Kenneth Waters takes a still more positive view. He sees different definitions of the molecular gene as unified by a 'fundamental gene concept', namely, 'a gene for a linear sequence in a product at some stage of genetic expression' (Waters, 1994, p. 178). Whether introns are parts of the gene, or whether transpliced sequences are parts of the same gene, depends on which particular 'linear sequence in a product at some stage of genetic expression' (ibid., p. 179) is the focus of the discussion in which the term 'gene' occurs. A gene, one might say, is the 'image' in the DNA of whatever molecule is the focus of interest. In some cases, such as those involving extensive mRNA editing, the resultant molecule has no such image, despite the fact that the precursor molecule derives from a DNA molecule. These molecules are 'gene products' for which there are no genes (Sarkar, 1996). Thomas Fogle is a less sanguine pluralist, arguing that current usage of the term 'gene' is driven by a historically-derived stereotype. This stereotype is based on facts about the structure and function of protein coding genes that take the form of a continuous series of DNA bases. More problematic DNA elements, with diverse functions and structures are called genes if they resemble the stereotype sufficiently. Sets of DNA elements that are discovered to underlie some function in the cell are divided into one or more genes and various auxiliary elements in order to facilitate seeing them via this stereotype (Fogle, 2001).

While Waters has looked for unity in the diversity of the gene concept, Lenny Moss has recently argued that one particular aspect of conceptual diversity is the key to understanding both the scientific utility of the gene concept and some of its pitfalls. According to Moss, both current and historic conceptualizations of the gene make use of two different ways of classifying DNA sequences, taxonomic schemes that he labels Gene-P and Gene-D:

> Gene-P is the expression of a kind of instrumental preformationism ... When one speaks of a gene in the sense of Gene-P one simply speaks *as if* it causes the phenotype. A gene for blue eyes is a Gene-P. What makes it count as a gene for blue eyes is not any definite molecular sequence (after all it is the absence of a sequence based resource that matters here) nor any knowledge of the developmental pathway that leads to blue eyes (to which the 'gene for blue eyes' makes a negligible contribution at most), but only the ability to track the transmission of this gene as a predictor of blue eyes. Thus far Gene-P sounds purely classical, that is, Mendelian as opposed to molecular. But a molecular entity can be treated as a Gene-P as well. BRCA1, the gene for breast cancer, is a Gene-P, as is the gene for cystic fibrosis, even though in both cases phenotypic probabilities based upon pedigrees have become supplanted by probabilities based upon molecular probes.
>
> ...

> Quite unlike Gene-P, *Gene-D is defined by its molecular sequence*. A Gene-D is a developmental resource ... which in itself is *indeterminate* with respect to phenotype ... To be a gene for N-CAM, the so-called 'neural cell adhesion molecule,' for example, is to contain the specific nucleic acid sequences from which any of 100 potentially different isoforms of the N-CAM protein may ultimately be derived ... N-CAM molecules are (despite the name) expressed at many tissues, at different developmental stages, and in many different forms. The phenotypes of which N-CAM molecules are co-constitutive are thus highly variable, contingent upon the larger context, and not germane to the status N-CAM as a Gene-D. So where a Gene-P is defined strictly on the basis of its instrumental utility in predicting a phenotypic outcome and is most often based upon the absence of some normal sequence, a Gene-D is a specific developmental resource, defined by its specific molecular sequence and thereby functional template capacity and yet it is indeterminate with respect to ultimate phenotypic outcomes. (Moss, 2001, pp. 87-88)

Moss's work falls into a 'developmentalist' tradition of criticism of overly simple conceptions of the role of genes in the construction of phenotypes (Oyama, Griffiths, & Gray, 2001). Developmentalists have often directed their criticisms at 'implicit preformationism'—the idea that phenotypic outcomes are preformed in a genetic cause as 'traitunculi' (Schaffner, 1998) rather than emerging epigenetically through the interaction of this and other causes.[4] Moss clarifies this criticism by arguing that the 'preformationism' embodied in the Gene-P conception is both a productive research strategy in its own epistemological sphere *and* the source of oversimplified and unhelpful ideas about the role of genes in development. It is a productive research strategy because genes really are statistical predictors of phenotypic outcome. But sequences that are identified as Gene-Ps immediately become a legitimate focus of interest as Gene-Ds. If the presence of a sequence is correlated with an outcome it is sensible to ask how that sequence contributes to development. This double life of the gene concept and of individual (token) DNA sequences makes it easy to misunderstand claims made about Gene-Ps and to hear them as if they were about Gene-Ds. Genes are conceptualized in terms of their contextual effects (Gene-P) and then treated as developmental causes *under that conceptualization*, rather than the Gene-D conceptualization appropriate for thinking about developmental questions. In this way, the effect that a Gene-D has on a phenotype in a particular developmental context comes to be treated as an intrinsic property of that Gene-D and as an inevitable consequence of its presence in an organism.

[4] The notion that genetic causation works by specific genes 'encoding' or being 'for' specific phenotypic traits is widespread outside biology and is highly problematic (Gannett, 1999). It is controversial, however, to what extent, if any, this way of thinking influences actual biological researchers. It is sometimes said that medical geneticists and/or behavior geneticists make use of this conception of genetic causation more than other biologists. Moss has argued that this way of thinking is far more widespread. It is just this sort of issue that the current study hopes to throw light upon.

The most popular notion of the gene in the wider community is undoubtedly the informational conception: a gene is a packet of developmental information, or an instruction for development. Like many other philosophers, we would argue that the notion of information deployed in this conception of the gene 'is little more than a metaphor that masquerades as a theoretical concept and . . . leads to a misleading picture of possible explanations in molecular biology' (Sarkar, 1996, p. 187). Developmental information is not stored in the literal genetic code, because the formal coding relation between DNA and polypeptides specifies only the primary structure of proteins (Godfrey-Smith, 1999; Griffiths, 2001). The other informational and cybernetic locutions used in molecular biology are equally applicable to genetic and non-genetic factors in development, and if taken seriously, require a recognition of the fact that the developmental significance of a causal factor is a function of its developmental context (Griffiths, 2001; Griffiths & Knight, 1998). In our view, the loose notion of the gene as a unit of developmental information does little harm when the focus of research is actual molecular gene products and their interactions and when Gene-P notions are being applied to that level of gene expression. At that level of analysis there are tight connections between what genes do (Gene-P) and their intrinsic molecular nature (Gene-D). Furthermore, contextual factors affecting whether genes are transcribed, what products are derived from them and what those products go on to do are the actual focus of research and unlikely to be systematically overlooked. When the focus is on higher levels of biological organization, however, and particularly when results are reported to the wider community, loose information talk almost inevitably leads to the conflation of Gene-P and Gene-D and the resultant misinterpretation discussed above (Griffiths, forthcoming; Moss, 2002).

For most purposes, evolutionary biologists work with something like the Gene-P conception. Their interest is in the relationship between changing gene frequencies in populations over time and changes in the phenotypes manifested by the individuals that make up those populations. The Gene-P conception embodies the relevant kind of statistical relationships. The term 'evolutionary gene concept', however, is normally attached to a very different idea introduced by George C. Williams (1966) and elaborated by Richard Dawkins: 'Any stretch of DNA, beginning and ending at arbitrarily chosen points on the chromosome, can be considered to be competing with allelomorphic stretches for the region of chromosome concerned' (Dawkins, 1982, p. 87). The purpose of the evolutionary gene concept is to abstract away from the complexities of the gene–phenotype relationship. The inheritance of DNA sequences is assumed to underlie all heritable phenomena of interest to evolution. Change over time in the DNA sequence can be exhaustively described using the formalism of population genetics and the evolutionary gene concept. That takes care of the molecular level, leaving the evolutionary biologist free to study the evolution of phenotypic characters using the formalism of population genetics without worrying about the relationship between molecular genes and these phenotypic characters. However, as many critics rapidly pointed out, the ability to describe changes in the composition of the genome after the fact is not the same as the ability to explain or predict those changes. Kim Sterelny and

Griffiths conclude that in his responses to these critics Dawkins effectively abandoned the evolutionary gene concept (Sterelny & Griffiths, 1999, pp. 79-82). The definition of the gene with which he replaced it with is an amalgam of Gene-P and Gene-D and has no particular currency in the biological community in its own right, so we have not considered it in this study. Williams has also moved on from his 1966 definition and now supports a very radical version of the informational conception of the gene (Williams, 1992).

On the basis of the literature reviewed above, we advanced three hypotheses:

Hypothesis 1. We expected to see a strong divergence between molecular and evolutionary biologists, given the emphasis on the investigation of the intrinsic, structural nature of the gene in the former discipline and the emphasis on genes as markers of phenotypic effects in the latter discipline. In particular, we expected molecular biologists to be reluctant to identify a gene only by its contributions to relatively distant levels of gene expression. Conversely, we expected evolutionary biologists to be reluctant to treat two similar DNA sequences as the same gene when they lead to different outcomes for the larger system in which they are embedded.

Hypothesis 2. We also expected developmental biologists and evolutionary biologists to differ, with evolutionary biologists emphasizing the predictive relationship between genes and phenotypes and developmental biologists emphasizing the intrinsic nature of the gene as a molecular object and contextual effects on gene expression. Consequently, we also expected stronger support for the informational conception of the gene from evolutionists.

Hypothesis 3. We expected developmental biologists to be less attracted to Moss's Gene-P and to the informational conception of the gene than (other) molecular biologists. We expected developmental biologists to be attracted to conceptions that emphasize contingency and context dependency, such as Moss's Gene-D and various developmentally-oriented conceptions of the gene canvassed in the literature on evolutionary developmental biology.

In addition to these specific hypotheses, which, as is evident, are particularly targeted at testing some of the ideas of Lenny Moss, we saw this as an exploratory study and were interested in what the responses suggest about the general state of the gene concept in contemporary biology. We also examined the effects of age and gender.

3. Methods

3.1. Subject recruitment

Subjects were post-Ph.D. biological scientists. We used the membership list of a campus-wide network of genomics and bioinformatics researchers at the University of Sydney and supplemented this list by examining the websites of the academic units in which list members were located. This produced a list of 250 potential subjects from biology, biochemistry, agriculture, veterinary science, medicine, pharmacology and chemistry. These scientists received a questionnaire and a covering

letter explaining in general terms the aim of the study. From this mail-out we received some eighty correctly completed responses, a reasonable response rate given that we had no prior contact with recipients.

3.2. Questionnaire design

The questionnaire had three sections, the first part designed to determine the subject's research field, the second asking them direct questions about the gene concept and the third asking them to apply the gene concept to specific cases. The first section of the questionnaire (Section A) gathered personal data on the professional training, research experience and current research field of subjects, along with age and gender (Table 1).

The five questions in Section B of the questionnaire (Table 2) contained direct questions about the definition of the gene, the function of the gene and the metho-

Table 1
Section A of the questionnaire. Used to group subjects by field, age and gender. Multiple selections were allowed.

1. Current Disciplinary Location	**2. Area of Training**
a. Medicine	a. Medicine
I. Molecular Medicine	I. Molecular Medicine
II. Oncology	II. Oncology
III. Biochemistry	III. Biochemistry
IV. Pharmacology	IV. Pharmacology
V. Infectious Diseases	V. Infectious Diseases
b. Biochemistry	b. Biochemistry
I. Molecular	I. Molecular
II. Cell	II. Cell
III. Protein	III. Protein
IV. Metabolism	IV. Metabolism
c. Biological Sciences	c. Biological Sciences
I. Genetics	I. Genetics
II. Development	II. Development
III. Evolution, Taxonomy	III. Evolution, Taxonomy
IV. Ecology	IV. Ecology
V. Microbiology	V. Microbiology
d. Agriculture	d. Agriculture
I. Agricultural Genetics	I. Agricultural Genetics
II. Animal Genetics	II. Animal Genetics
III. Animal Science	III. Animal Science
IV. Plant Breeding	IV. Plant Breeding
e. Veterinary Science	e. Veterinary Science
f. Pharmacology	f. Pharmacology
g. Other	g. Other

3. Disciplines of Degree	**4.a. Gender**	**b. Age**
Undergraduate degree in...	1. female	1. 20–34
Postgraduate degree other than Ph.D. in...	2. male	2. 35–49
Ph.D. defended in...		3. 50–70

Table 2
Section B of the questionnaire. In the free choice task, subjects were asked to tick every acceptable answer alternative. In the forced choice task they were asked to choose the best alternative. The 'Other' alternative for each question has been omitted.

1. In short: What is a gene?

1. That which makes the difference between two phenotypes.
2. Any Nucleic acid sequence whatsoever.
3. Nucleic acid sequence with a certain characteristic structure.
4. Nucleic acid sequence with a certain characteristic function.
5. A carrier of heritable information.
6. A resource for Development.

2. What is the biological function of a gene?

1. Causing a phenotypic outcome.
2. Determining a phenotypic outcome.
3. Coding for the primary structure of a protein.
4. Providing a developmental resource, on a par with other (epigenetic and environmental) resources, for the construction of the organism.
5. To channel and reinforce epigenetic propensities instead of specifying incremental alterations in morphology.
6. Releasing and biasing the expression of latent morphogenetic capacities.
7. Mechanism to buffer the development of established and ecologically successful phenotypes against environmental perturbations and metabolic noise.
8. Functional part of a program for development that is written in the sequence of nucleotide bases.

3. What makes two genes "homologous"?

1. Both have derived from a common ancestral gene (they are orthologous).
2. Both have an identical sequence of nucleotides.
3. Both produce functionally equivalent molecular products.
4. Both are situated at homologous sites on homologous chromosomes.
5. Both are able to recombine with one another in practice.
6. Both are able to recombine with one another in theory (physically compatible).
7. Both have derived from a gene duplication (they are paralogous).

4. What is the methodological value of the gene concept?

1. A gene has instrumental utility in predicting a phenotypic outcome.
2. Central concept in evolution: allows i) shortcut definition of evolution as change in gene frequency and ii) a general conception of evolution as gene selection.
3. Studying the biological role of a particular gene, which involves locating it within the contexts in which it is biologically active, helps to elucidate the complex molecular pathways in which it is an interactant.

4. A convenient entry point to functionally conserved multi-molecular modules as units of development, morphology, variation and innovation.
5. 'Gene' functions to remind modern geneticists of what it is that make a region of nucleic acid 'interesting', or of what constitutes 'meaningful structure' in the genome.
6. A gene draws our attention to a collection of useful functional domains (exons) which can be combined in different ways.
7. A handy and versatile term whose meaning is determined by the context in which it is used.

5. At length: What is a gene:
1. The functional and physical unit of heredity passed from parent to offspring.
2. A stretch of DNA sequence that codes for a particular protein that has a particular function.
3. A package of information that contains and implements a particular instruction.
4. A gene is defined by its relationship to a phenotype regardless of the specific molecular sequence and the whole developmental mechanisms involved.
5. A developmental resource defined by its specific molecular sequence and functional template capacity but which is indeterminate with respect to the phenotypic outcomes to which it will contribute.
6. A segment of chromosome. Some genes direct the synthesis of proteins, others have regulatory functions.
7. A process that includes DNA sequences and other components, which participate in the time and tissue specific expression of a particular polypeptide product.
8. Any stretch of DNA, beginning and ending at arbitrarily chosen points on the chromosome, that segregates and recombines with appreciable frequency.
9. A functional unit and part of the processes that specify cellular and intercellular organization, defined by the action of a complex self-regulating system for which the inherited DNA provides the crucial raw material.

dological value of the gene concept. The answer alternatives for each question were designed to capture the various conceptions of the gene discussed in the literature. We used a number of different formulations of each conception to avoid superficial effects, such as antipathy to particular words or phrases. The actual wordings of many of the answer alternatives were copied, or lightly adapted, from the literature and from genomics websites, so as to avoid claims that we have misrepresented the relevant ideas, or attempted to influence responses through biased formulations.[5] Each question had an 'Other' alternative in which subjects could supply their own answer, but no useful data was obtained by this means.

This section of the questionnaire contained both 'free choice' and 'forced choice' tasks. The former required subjects to indicate for each question all the answer alternatives to which they could agree. The latter required subjects to choose the single best answer amongst the alternatives offered. The free choice task was designed to recognize that individual subjects could conceptualize genes in more than one way. The forced choice task aimed at revealing some of the subjects' preferences and helped to increase variance when the free choice responses failed to exhibit significant differences between the subject groups.

Section C of the questionnaire (Table 4) was based on the design of an informal study conducted by Rob C. Knight in New Zealand with 10 respondents. This section used 'indirect' questions, asking subjects to apply their conception of the gene, rather than to answer questions about it. Subjects were given twenty-two examples of specific ways in which two DNA sequences could differ from one another and asked whether, in each case, these DNA sequences were two copies of the same gene.

3.3. *Defining groups*

All three hypotheses assume the existence of more or less clearly identifiable groups of biologists that differ significantly in their views about genes. Using responses to Section A of the questionnaire (Table 1) we defined the three groups of biologists that appear in our hypotheses—'molecular', 'developmental', and 'evolutionary'' A fourth group of 'whole organism biologists' failed to show any significant differences and is not further discussed. Our operational definitions of molecular, developmental, and evolutionary biologists were Boolean combinations of answers to questions in Section A, based on our intuition about the kind of scientist one can expect to find in certain (sub–)disciplines, departments, and research fields.[6] Essentially, anyone who crossed 'developmental biology' either as area of training or Ph.D. or disciplinary location/current research was classified as a member of the 'developmental' group, and the 'molecular' group had the majority of their training and work in biochemistry or molecular biology, while an

[5] A fully annotated version of the questionnaire indicating these sources is available in the documents linked to this paper on the Philosophy of Science Association preprint server (http://philsci-archive.pitt.edu).

[6] These definitions are available in the documents linked to this paper on the Philosophy of Science Association preprint server (http://philsci-archive.pitt.edu).

'evolutionist' needed to tick evolution or taxonomy or ecology as either current location or training. Predictably, these definitions created overlapping groups. Because we were concerned to maintain an adequate number of subjects, those in the intersection of two groups were examined on an individual basis and assigned to a single group by a subjective assessment of their overall pattern of responses to Section A. This process was completed before examining Sections B and C of the questionnaire in order to keep group membership independent of subjects' views on the gene concept.

3.4. Data analysis

The sample sizes of some of our groups (notably 'evolutionary') were small. This does not invalidate the results reported below, since the statistical measures utilized are suitable for small sample sizes and unequal group sizes. This notwithstanding, results based on such small sample sizes must be taken as suggestive rather than definitive, and as a basis for further studies.

For our analysis we used tests suitable for categorical data, which are all based on cross-tabulation. We used the chi-square ($\chi 2$) for determining the *presence* of an association between our independent variables (kinds of biologists) and dependent variables (different conceptions of the gene). Our measure of *strength* of association in many of our free choice cases (which can be represented by a two-way contingency table) was the phi coefficient (Φ). In such symmetric cases and a high enough sample size it mimics the correlation coefficient by having a maximum value of 1 (perfect correlation) and a minimum value of 0 (no association), and Phi can be interpreted as a symmetric version of percent difference.

Many of our contingency tables, however, involved variables with more than two values, (for example, the forced choice tasks), and small sample sizes. In these cases Phi can be infinitely larger than 1 (which does not lend itself to an easy interpretation of the test in terms of the strength of association), and we preferred another statistical test designed for groups of unequal size and small sample sizes, Cramer's V, which gives good norming from 0 to 1 regardless of table and sample size. It is worth noting that since Phi as well as Cramer's V are symmetrical measures, they tend to understate *asymmetric* relationships between the independent and dependent variable. Also, the more unequal the marginals, the more V will understate an association.

With respect to the significance of the results we have followed convention by dividing our results into those significant at the 0.05 (5%) level, those significant at the 0.1 (10%) level and reported other associations as not significant ('ns'). Failure to achieve the desired level of significance may often reflect the lack of power in this study due to small sample sizes. For the same reason, the *absence* of an association in the results tables below does not provide good evidence that those variables are independent of one another.

4. Results

4.1. The current state of the gene concept

The responses suggest that the classical molecular gene concept continues to act as an important point of departure for biologists in conceptualizing the gene. When subjects were asked to indicate the biological function of a gene, the answer alternative corresponding to the classical molecular conceptualisation was overwhelmingly the most popular: *2.3. Coding for the primary structure of a protein* (free choice 92%, forced choice 63%) (italicized phrases are drawn from the questionnaire. See Table 2 for a full list of questions and answers). Similarly, when subjects were asked to choose between a series of sentence-length definitions of the gene, the answer alternative corresponding to the classical molecular conceptualization was the second most popular: *5.2. A stretch of DNA sequence that codes for a particular protein and that has a particular function* (free choice 89%, forced choice 24%). This was the definition offered by the (Australian) National Human Genome Research Institute website, so the high level of agreement is perhaps unsurprising. However, the most popular answer to this question was a very broad, Mendelian (or even pre-Mendelian!) definition: *5.1. The functional and physical unit of heredity passed from parent to offspring* (free choice 89%, forced choice 43%). The significance of this finding is explained in the next section.

Amongst a set of shorter phrases purporting to sum up what a gene is, an alternative in the spirit of the classical molecular gene concept was again the second most popular choice: *1.4. Nucleic acid sequence with a certain characteristic function* (free choice 79%, forced choice 35%). The most popular of these short phrases, however, was *1.5 Carrier of heritable information* (free choice 87%, forced choice 44%). So, at least as a shorthand, the informational conception of the gene has currency amongst working biologists.

In response to the more demanding question, 'What is the methodological value of the gene concept?' the most popular answer alternative was *4.3. Studying the biological role of a particular gene, which involves locating it within the contexts in which it is biologically active, helps to elucidate the complex molecular pathways in which it is an interactant* (free choice 77%, forced choice 36%). This suggests that the complexities of genetic causation are salient facts for working biologists.

4.2. Age and gender results

Subjects were divided into three age groups: 20–34, 35–49 and 50–70 and into male and female. Several answers were correlated with gender, but all these associations disappeared when the association between age and gender was taken into account. Female subjects were on average much younger than male subjects. The 20–34 age group contained 33.3% of the female subjects, but only 14% of the male subjects; the 35–49 age group contained 54.2% of females and 40.4% of males; the 50–70 age group contained only 10.3% of females, but 45.6% of males.

When asked about the methodological value of the gene concept, older subjects favored two strikingly pluralistic, deflationary statements: *4.5. 'Gene' functions to*

remind modern geneticists of what it is that make a region of nucleic acid 'interesting', or of what constitutes 'meaningful structure' in the genome and *4.7. A handy and versatile term whose meaning is determined by the context in which it is used.* Older subjects were also more likely to accept a third statement that located the gene primarily in evolutionary biology (*4.2. Central concept in evolution: allows i) shortcut definition of evolution as change in gene frequency and ii) a general conception of evolution as gene selection*). In contrast, younger biologists overwhelmingly saw the gene as primarily an object of interest to molecular biology (answer 4.3, quoted above). Results are shown in Fig. 1.

We have already noted that when asked to choose between a series of sentence-length definitions of the gene, subjects predominantly favored either the classical molecular gene concept in a formulation obtained from Australia's National Human Genome Research Institute (5.2) or a very vague, Mendelian (or even pre-Mendelian!) definition of the gene as unit of heredity (5.1). These two definitions turn out to be extremely strongly associated with age, younger subjects favoring the 'molecular' definition and older subjects the '(pre-) Mendelian' definition (Fig. 2).

The fact that older biologists are more inclined to assign the gene a role in evolutionary theory is probably due to the fact that biologists with an evolutionary focus were concentrated in our oldest age group, which in turn results from the fact that departments with an evolutionary focus have been growing (if at all) much more slowly than departments with a molecular focus in the last few decades. The finding that older biologists take a more pluralistic view of the gene cannot be dismissed in the same way. Pluralism, we suspect, is a genuine function of age, which is perhaps to be expected, given the regularity with which cherished ideas about the gene have been overthrown in the last fifty years.

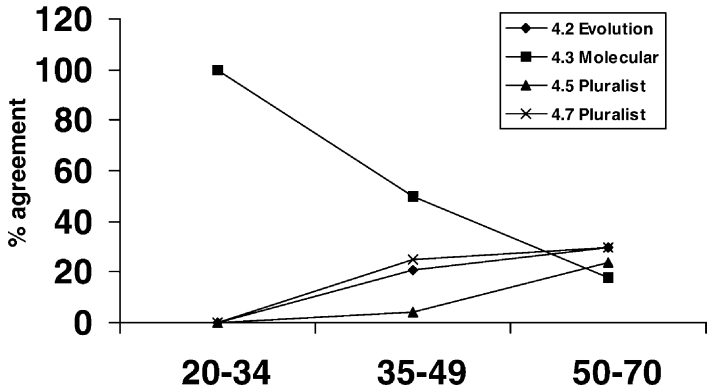

Fig. 1. Age related results for the forced choice task on Section B, Question 4. See Table 2 and text for details. Association 0.625, significance 0.012.

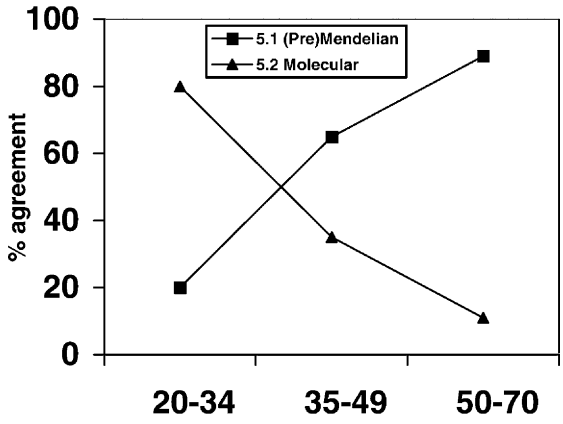

Fig. 2. Age related results for the forced choice task on Section B, Question 4. See Table 2 and text for details. Association 0.590, significance 0.009.

4.3. Results for Hypothesis 1

Hypothesis 1 predicts a strong divergence between molecular and evolutionary biologists, given the emphasis on the investigation of the intrinsic, structural nature of the gene in the former discipline and the emphasis on genes as markers of phenotypic effects in the later discipline. In particular, we expected molecular biologists to be reluctant to identify a gene only by its contributions to relatively distant levels of gene expression. Conversely, we expected evolutionary biologists to be reluctant to treat two similar DNA sequences as the same gene when they lead to different outcomes for the larger system in which they are embedded.

In the light of this hypothesis we made the following predictions for the free choice task on Section B of the questionnaire (Table 2):

Question 1. The molecular group is more likely to accept a structural conception of the gene (answer alternative 1.3) than the evolutionary group. The evolutionary group is more likely to accept the Gene-P option (1.1) than the molecular group.

Question 2. The molecular group is more likely to accept the classical molecular conception (2.3). The evolutionary group is more likely to accept the Gene-P variants 2.1 v 2.2. (Because this was a free choice task, when more than one answer alternative seemed to express the same conception of the gene, we predicted a weak disjunction of these alternatives—either a or b or both—as the response for the relevant group.)

Question 4. The evolutionary group is more likely to accept the Gene-P alternative 4.1 and the specifically evolution-oriented 4.2. The molecular group is more likely to accept 4.3 and 4.6, answers that emphasize the investigation of genes at the molecular level.

Question 5. Evolutionary biologists are more likely to accept the gene-P option 5.4, whilst the molecular group is more likely to accept statements of the classical molecular conception (5.2 v 5.6) and the more complex answers relating genes to other processes at the molecular level (5.5 v 5.9).

The results for the free choice task were not consistent with Hypothesis 1. Rather than favoring the phenotype-focused Gene-P conception, the evolutionary group tended to *reject* it when it was offered explicitly (Table 3). The forced choice results were similar, and are not reported here for reasons of space (complete results are available at http://philsci-archive.pitt.edu). So the results from Section B, in which we asked direct questions, suggest that biologists with an evolutionary focus in their research do not conceptualize genes in terms of their phenotypic effects in any way that distinguishes them from biologists with a purely molecular research focus. A very different picture emerges, however, from responses to the indirect questions in Section C (Table 4).

In the light of Hypothesis 1, we predicted that the molecular and evolutionary groups would respond differently to several items in Section C of the questionnaire (see Table 4). Questions 6.4, 6.5, 6.6 and 6.7 present a series of cases in which the proposed identity between two DNA sequences is based on their producing identical proximal gene products, whilst their more distal products, with more direct causal relevance to the phenotype, are increasingly allowed to diverge. We expected the molecular group to classify all these pairs as the same gene, because of their proximal similarity. Whilst both groups should agree that 6.4 represents two copies of the same gene, we expected the evolutionary group to reject the claim that the sequences in 6.5, 6.6 and 6.7 are two copies of the same gene, because the distal products of the two sequences differ in ways that should lead to different phenotypic effects. Questions 6.9 and 6.10 also describe sequences which have the same nucleotide sequence but different molecular products, and so we expected the molecular group to treat these as the same gene and the evolutionary group to treat them as different genes. Conversely, 6.19 describes sequences that differ at the molecular level but have the same effect on the phenotype, a straightforward instance

Table 3
Test of Hypothesis 1 with data from the free choice task. Left-hand column shows answer alternatives for which we predicted agreement by the molecular group (M), right-hand column those for which we predicted agreement by the evolutionary group (E). Results: the numbers behind the characters show percentage of yes answers among the respective group (M, E), the following fractions indicate strength (0–1) and significance (0–1) of association. Results marked ns were not significant (>10% or .100). **Bold** results indicate high significance (<5% or .050), *italic* results show associations in the reverse direction to that predicted.

1.3	M 50%, E 50%, ns	1.1	*E 0%, M 43%, ns*
2.3	M 100%, E 100%, ns	2.1v2	***E 0%, M 71%, .458/.002***
4.3	M 77%, E 80%, ns	4.1	*E 40%, M 72%, ns*
4.6	M 39%, E 40%, ns	4.2	**E 100%, M 46%, .342/.023**
5.2v6	M 95%, E 100%, ns	5.4	E 20%, M 10%, ns
5.5v9	M 55%, E 60%, ns		

Table 4
Section C of the questionnaire. Subjects were asked whether each item described two copies of the same gene. Question 6.8 contained a typological error.

6.1 Any two identical nucleotide sequences, beginning and ending at arbitrary points, at equivalent loci on homologous chromosomes in different cells of the same organism.
6.2 Two transcription units of identical nucleotide sequence at equivalent loci on homologous chromosomes in different cells of the same organism.
6.3 Two transcription units of identical nucleotide sequence on non-homologous chromosomes in the same organism.
6.4 Two transcription units of identical nucleotide sequence leading to the same functional protein.
6.5 Two transcription units of identical nucleotide sequence, which are translated into the same polypeptide chain, regardless of how it is fo folded.
6.6 Two transcription units of identical nucleotide sequence, which produce the same final transcript, regardless of what happens to this transcript.
6.7 Two transcription units of identical nucleotide sequence, which produce the same primary transcript, regardless of what happens to this transcript.
6.9 Two transcription units of identical nucleotide sequence whose final transcript contains differently spliced exons.
6.10 Two transcription units of identical nucleotide sequence, one of which has its exons scrambled in its final transcript (as happens in ciliates).
6.11 Two transcription units of identical nucleotide sequence with different promoters but with identical levels of transcription.
6.12 Two transcription units of identical nucleotide sequence with different promoters and different levels of transcription.
6.13 Two transcription units which differ only in a single silent mutation.
6.14 Two transcription units that differ in a number of silent mutations, not affecting the level of expression.
6.15 Two transcription units that differ in a number of silent mutations, affecting the level of expression significantly.
6.16 Two otherwise identical transcription units containing different nonsense mutations both of which destroy the corresponding enzyme's catalytic activity.
6.17 Two transcription units, which differ so as to produce a single substitution in the amino acid sequence but with no observable developmental effect.
6.18 Two transcription units with identical sequences but which produce different polypeptides due to differences in the genetic code (eg., between mitochondria and nuclei).
6.19 Two allelic transcription units differing in sequence but with identical phenotypic effect.
6.20 Two transcription units of identical nucleotide sequence, one of which is found on a free transposon and one of which is found in normal genomic DNA.
6.21 Two identical nucleotide sequences, one is an active coding sequence, the other is split into two (non-functional) pieces by an insertiiiiition. ion.
6.22 Two transcription units of identical nucleotide sequence that have evolved independently in different taxa through convergent evolution.

Table 5
Test of Hypothesis 1 with data from the indirect section. In this table we expected for all but the last answer alternative (6.19) to find a higher level of agreement among molecular biologists. Result cells: the numbers behind the characters show percentage of yes answers among the respective group (M, E), the following fractions indicate strength (0–1) and significance (0–1) of association. Results marked 'ns' were not significant (>10% or .100). **Bold** results indicate high significance (< 5% or .050).

Answer predictions for Molecular Group			
6.5	**M 66%, E 0%, .389/.006**	6.6	**M 66%, E 20%, .285/.046**
6.7	**M 60%, E 0%, .358/.012**	6.9	**M 50%, E 0%, .304/.033**
6.10	M 35%, E 0%, ns		
Prediction for Evolutionary Group	E 20%, M 16%, ns		
6.19			

of same Gene-P, and we expected the evolutionary group to accept this as a case of the same gene and the molecular group to reject it.

In contrast to the results for the direct questions in Section B, those for the indirect questions in Section C supported Hypothesis 1 (Table 5). The pattern of answers to questions 6.5, 6.6 and 6.7 shows the evolutionary group responding significantly more strongly to changes in distal function than the molecular group. Answers to 6.10 and 6.19 differed in the predicted direction, although the difference was not statistically significant. However, the fact that in the direct questions the evolutionary group was *less* accepting of Gene-P conceptions than the molecular group and that this position is here apparently reversed, leads us to take these insignificant results as at least somewhat suggestive. The striking difference between the results for Section B and Section C of the questionnaire lead us to advance a new hypothesis, namely that the evolutionary group has an explicit belief that genes are molecular entities and should be defined and investigated at that level, but that when asked to think about actual cases, they employ a Gene-P conception that abstracts away from differences at the molecular level and focuses on phenotypic effects. We hope to investigate this hypothesis in future research.

4.4. Results for Hypothesis 2

Hypothesis 2 predicts that evolutionary biologists will emphasize Gene-P and effects on the phenotype, whilst developmental biologists emphasize Gene-D and contextual effects on gene expression. Consequently, we also expected stronger support for the informational conception of the gene from evolutionists.

This hypothesis predicts that the developmental group will differ from the evolutionary group in a fairly similar way to the molecular group. We therefore predicted a similar pattern of responses for these groups on the free choice task on Section B of the questionnaire (see Table 2) as we predicted for the molecular and evolutionary groups when assessing Hypothesis 1 in the last section. The main difference is the addition of a preference for some developmental and contextual notions of the gene (1.6, 2.4, 2.5, 2.6, 2.7, 5.7) on the part of the developmental group. The predictions and results are given in Table 6.

Table 6
Test of Hypothesis 2 with data from the free choice task. The column on the left show answer alternatives for which a higher level of agreement was predicted for the developmental group (D). The column on the right shows our expectations for the evolutionary group (E). Results: the numbers behind the characters show percentage of yes answers among the respective group (D, E), the following two fractions indicate strength (0–1) and significance (0–1) of association. 'Ns' indicates that the result was not significant (>10% or .100). **Bold** results indicate high significance (< 5% or .050), *italics* highlight outcomes reverse from the stated prediction.

1.3	D50%, E 50%, ns	1.1	*E 0%, D 25%, ns*
1.6	D 50%, E 20%, ns	2.1v2	***E 0%, D63%, .625l.024***
2.3	*D 75%, E 100%, ns*	4.1	*E 40%, D 75%, ns*
2.4–7	D 75%, E 40%, .350/.207, ns	4.2	E 100%, D 75%, ns
4.3	D 88%, E 80%, ns	5.4	E 20%, D 0%, ns
4.6	D 38%, E 40%, ns		
5.2v6	*D 88%, E 100%, ns*		
5.5v9	D 75%, E 60%, ns		
5.7	D 38%, E 0%, .433/.118, ns		

The results for the free choice task were not consistent with Hypothesis 2. Rather than favoring the phenotype-focused Gene P conception of the gene, the evolutionary group tended to *reject* it when it was offered explicitly (Table 6). So the results from Section B, in which we asked direct questions, suggest that biologists with an evolutionary focus in their research are as exclusively focused on the molecular-level properties of the gene as biologists with a developmental research focus, if not more so. But, just as with Hypothesis 1, a different picture emerges from the indirect questions in Section C.

In the light of Hypothesis 2, we predicted that the developmental and evolutionary groups would respond differently to several items in Section C of the questionnaire (Table 4). As already noted, Hypothesis 2 predicts that the developmental group will differ from the evolutionary group in a similar way to the molecular group. We therefore predicted similar responses for these groups on the forced-choice task on Section C of the questionnaire as we predicted for the molecular and evolutionary groups when assessing Hypothesis 1. The predictions and results are given in Table 7.

In contrast to the results for the direct questions, and just as we saw with Hypothesis 1, the results for the indirect questions supported Hypothesis 2 (Table 7). The pattern of answers to questions 6.5, 6.6 and 6.7 shows the evolutionary group responding more strongly to changes in distal function than the developmental group, although the results for the first two questions fail to reach the 5% significance level. The results for 6.9 and 6.10 show a significant association in the predicted direction (the corresponding results for Hypothesis 1 showed an association in the predicted direction, but did not achieve a 5% significance level, which is plausibly because the test lacked power). The contrast between the results for direct questions (Section B) and indirect questions (Section C) leads us to repeat our tentative suggestion above that the evolutionary group has an explicit belief that genes are molecular entities and should be defined and investigated at that

Table 7
Test of Hypothesis 2 with data from the indirect section. In this table we listed the indirect questions (Table 4) for which we expected a higher level of agreement from the developmental group when compared to the evolutionary. Result cells: the numbers behind the characters show percentage of yes answers among the respective group (D, E), the following fractions indicate strength (0–1) and significance (0–1) of association. Results marked 'ns' were not significant (>10% or .100). **Bold** results indicate high significance (< 5% or .050).

Predicted answers for Developmental group			
6.5	D 33%, E 0%, ns	6.6	D 67%, E 20%, .447/.094
6.7	**D 56%, E 0%, .556/.038**	6.9	**D 67%, E 0%, .645/.016**
6.10	**D 56%, E 0%, .556/.038**		

level whilst deploying in their actual thinking a Gene-P conception that abstracts away from differences at the molecular level and focuses on phenotypic effects.

4.5. Results for Hypothesis 3

Hypothesis 3 predicts that developmental biologists will be less attracted to Moss's Gene-P and to informational conceptions of the gene than (other) molecular biologists. We expected developmental biologists to be more attracted to conceptions that emphasize contingency and context dependency, such as Moss's Gene-D and various developmentally oriented definitions of the gene which we quoted from the literature on evolutionary developmental biology.

In the light of this hypothesis we made the following predictions for the free choice task on Section B of the questionnaire (see Table 2):

> Question 1. The molecular group will be more likely to accept the Gene-P option (1.1) than the developmental group, who will be more likely to accept the Gene-D option (1.6).
> Question 2. The developmental group will be more likely to accept some of the more or less radical epigenetic options (2.4–2.7). The molecular group will be more likely to accept the Gene-P variants 2.1 and 2.2.
> Question 4. The molecular group will be more likely to accept the Gene-P alternative 4.1, while the developmental group will be more likely to accept 4.3 and 4.6, answers that emphasize the investigation of genes at the molecular level, and also 4.4, an option highlighting the idea of developmental modularity.
> Question 5. Molecular biologists will be more likely to accept the Gene-P option 5.4, while the developmental group should like the Gene-D version given by 5.5

Results are shown in Table 8.

The results were broadly supportive of the hypothesis. Alternatives 1.6 and 5.5, which we saw as embodying the Gene-D conception, discriminated strongly between the two groups, with the association for 5.5 being significant at the 5% level. Question 5.5 could also be regarded as testing for interest in the context sensitivity of gene expression. Alternatives 1.1 and 2.1, which we saw as embodying the Gene-P conception, also discriminated strongly between the two groups, and

Table 8
Test of Hypothesis 3 with data from the free choice task. Left-hand column shows answer alternatives for which we predicted agreement by the developmental group (D), right-hand column those for which we predicted agreement by the molecular group (M). Result cells: the numbers behind the characters show percentage of yes answers among the respective group (D, M), the following fractions indicate strength (from 0–1) and significance (0–1) of association. Results marked 'ns' were not significant (>10% or .100). **Bold** results indicate high significance (< 5% or .050), *italic* results show associations in the reverse direction to that predicted.

1.6	D 50%, M 21%, ns	1.1	M 43%, D 14%, ns
2.4–7	D 60%, M 53%, ns	2.1	**M 41%, D 0%, .353/.041**
4.3	D 88%, M 75%, ns	2.2	M 72%, E 60%, ns
4.4v6	D 60%, M 55%, ns	4.1	M 73%, D 60%, ns
5.5	**D 40%, M 13%, .282/.048**	5.4	M 20%, D 0%, ns

the latter association was significant at the 5% level. The stronger preference for the Gene-D conceptualization in the molecular group is particularly striking, since according to hypotheses one and two both these groups should show some tendency to prefer a Gene-D to a Gene-P conceptualization, at least as compared to the evolutionary group.

We used the same five questions from Section B for a further test of Hypothesis 3 by considering the answers to those questions on the forced choice task, rather than the free choice task. The purpose of the forced choice task was to reveal differences hidden by the free choice task, in which minimally acceptable options may not be distinguished from highly preferred options. Just as with the free choice task we predicted the answers that we expected from each group for each question. Because this was a forced choice task in which each subject chose only one option, where more than one answer option seemed equally likely to be preferred by a particular group, we coded a strong disjunction of these answers as a single answer. Table. 9 shows the result of our grouping and recoding exercise for the forced choice answers according to Hypothesis 3.

As expected, the forced choice task discriminates more strongly between the two groups under scrutiny. Only Question 4 did not discriminate, the rest showing strong associations (between .430 and .630) with good significance (= 1%).

In their responses to Question 1, the molecular group showed a preference for the Gene-P and informational gene options (1.1 v 1.5), while the developmental

Table. 9
Grouped forces choice predictions for Hypothesis 3. The answer alternatives in each cell were combined by strong disjunction on the grounds of their expected appeal to one group.

Predictions for Developmental Group	**Predictions for Molecular Group**
1.3, 1.4, 1.6	1.1, 1.5
2.4, 2.5, 2.6, 2.7, 2.8	2.1, 2.2, 2.3
4.3, 4.4, 4.6	4.1
5.5, 5.7, 5.9	5.2, 5.3, 5.4, 5.6

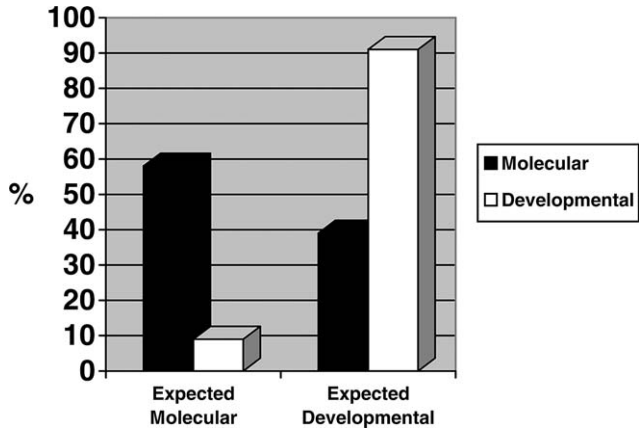

Fig. 3. Results for Hypothesis 3 for the forced choice task on Section B, Question 1. Association 0.430, significance 0.011.

group was distributed between the classical molecular gene options and the Gene-D option (1.3 v 1.4 v1.6). Results are shown in Fig. 3.

In response to Question 2, the molecular group preferred Gene-P conceptions and the classical molecular conception (2.1 v 2.2 v 2.3), whereas more than half of the developmental group voted for one of the developmental or epigenetic alternatives (2.4 v 2.6 v 2.7 v 2.8) (Fig. 4).

For Question Five we predicted that the developmental group would be more likely to choose the alternatives that emphasize the role of genes in a larger developmental context (5.5 v 5.7 v 5.9), while the molecular group would be more likely

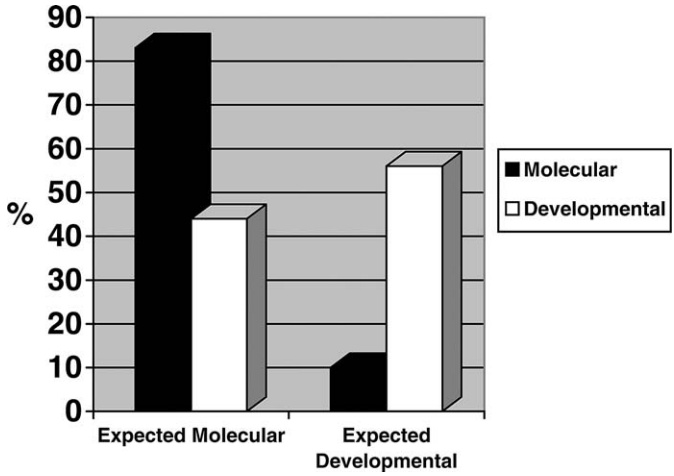

Fig. 4. Results for Hypothesis 3 for the forced choice task on Section B, Question 2. Association 0.460, significance 0.006.

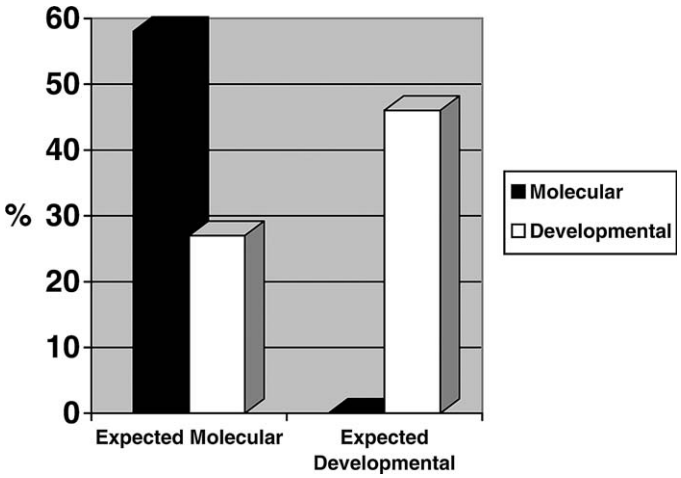

Fig. 5. Results for Hypothesis 3 for the forced choice task on Section B, Question 5. Association 0.630, significance 0.000.

to choose Gene-P and informational conceptions or alternatives reflecting the classical molecular conception (5.3 v 5.2 v 5.4 v 5.6). This pattern was in fact observed (Fig. 5), and closer examination revealed that none of the molecular group chose any of the developmental alternatives.

The indirect questions in Section C of the questionnaire (see Table 4) provided another test of Hypothesis 3. We did not expect as great a difference between the molecular and developmental groups as between either of these and the evolutionary group. Nevertheless, since Hypothesis 3 attributes to the molecular group a tendency to be more focused on the intrinsic similarity between DNA sequences themselves and less focused on similarities or differences in context, we expected the molecular group to be more likely to accept as two copies of one gene the pairs of sequences described in questions 6.5, 6.6 and 6.7, questions that introduce differences at progressively more distant stages of gene expression. Questions 6.18 and 6.20 also tested for whether sequence similarity would overwhelm other considerations. For question 6.19, however, although we had elsewhere predicted that molecular and developmental groups would be less attracted to Gene-P conceptions than evolutionary biologists, we predicted that the developmental group would be *most* unlikely to accept a judgment based on this conception, since it abstracts away from precisely the issues on which their research is focused. We therefore predicted that they would be less likely to accept the case described in 6.19 than the molecular group. The associations we actually observed for these various questions were all in the directions predicted, but only two of them were significant (Table 10).

Table 10
Test of Hypothesis 3 with data from section C. In this table we listed the indirect questions (Table 4) for which we expected a higher level of agreement from the molecular biologists when compared to the developmental group. Result cells: the numbers behind the characters show percentage of yes answers among the respective group (M, D), the following fractions indicate strength (0–1) and significance (0–1) of association. Results marked 'ns' were not significant (>10% or .100). **Bold** results indicate high significance (< 5% or .050).

Answer	Prediction for Molecular Group		
6.5	**M 66%, D 27%, .319/ .021**	6.18	M 20%, D 0%, ns
6.6	M 69%, D 54%, ns	6.19	M 16%, D 0%, ns
6.7	M 61%, D 46%, ns	6.20	**M 46%, D 10%, .313/ .024**

5. Conclusions and prospects for future research

The results reported here provide tentative support for all three hypotheses. Hypothesis 3 seems the most strongly supported. Biologists whose research focus is in developmental biology seem to conceptualize genes in a distinctive way, a way that appears to reflect their use of the gene concept to investigate the complex, developmental pathways through which genes are expressed. Hypotheses one and two, which suggest, in broad terms, that biologists whose research focus is in evolutionary biology conceptualize genes primarily via their effects on phenotypes, are supported in some tests but not others. The fact that the hypotheses are supported when indirect questions are used, but not when direct questions are used, leads us to advance an intriguing further hypothesis. We propose that these biologists may have an explicit belief that genes are molecular entities and should be defined and investigated at that level whilst deploying in their actual thinking about genetic problems a conception of the gene that abstracts away from differences at the molecular level and focuses on phenotypic effects.[7] We hope to test this hypothesis in future research.

Our general results for the whole subject population are consistent with Fogle's suggestion that the classical molecular gene concept continues to function as something like a stereotype for biologists, despite the many cases in which that conception does not give a principled answer to the question of whether a particular sequence is a gene (Fogle, 2001). Given the extensive psychological literature on this kind of cognitive structure and on the reasoning processes it supports, this also suggests productive lines of future inquiry.

Given the small number of subjects in this study and the simple criteria used to group them for statistical analysis, we are very encouraged by the ability of the study to discern differences between the groups. In ongoing research in the United

[7] Raphael Falk (personal communication) has suggested that this may be a general model for thinking about 'gene concepts'. Biologists across all fields accept that there is a material basis for heredity, with DNA molecules at its core, but their different structural definition of a gene (explicit or implicit) reflects which functions of DNA elements are relevant to research questions in specific fields.

States we are attempting to increase the number of subjects by an order of magnitude, and to use more sensitive measures to define our groups, including the techniques subjects utilize in their research, their individual ranking of journals and their attendance at professional meetings. Our results clearly indicate the importance of distinguishing between explicit and implicit ideas about the gene. In our ongoing research we ask subjects to engage in tasks such as dividing a limited research budget between a number of proposed research projects or indicating their confidence that a result will extrapolate from one model system to another. These tasks have the added advantage of providing numerical rather than categorical date, allowing a wider range of statistical procedures to be employed.

Acknowledgements

This research was supported by Australian Research Council Large Grant A-59906145. During the preparation of the paper Karola Stotz was supported by a Steven Manners Development Grant from the University Center for Social and Urban Research, University of Pittsburgh. The full data sets derived from the study are available in the documents linked to this paper on the Philosophy of Science Association preprint server (http://philsci-archive.pitt.edu). Part of the research instrument used in this study (Part C, Table 4) derived from a questionnaire designed by Rob Knight at the University of Otago, Dunedin, NZ, 1998. In preparing the paper we have benefited from the comments of Alan Love and Raphael Falk.

References

Dawkins, R. (1982). *The extended phenotype: The long reach of the gene*. San Francisco: Freeman.

Epp, C. D. (1997). Definition of a gene. *Nature*, *389(9 October)*, 537.

Falk, R. (1986). What is a gene? *Studies in History and Philosophy of Science*, *17*, 133–173.

Fogle, T. (2001). The dissolution of protein coding genes in molecular biology. In R. F. Beurton, R. Falk, & H.-J. Rheinberger (Eds.), *The concept of the gene in development and evolution* (pp. 3–25). Cambridge: Cambridge University Press.

Gannett, L. (1999). What's the cause? The pragmatic dimensions of genetic explanation. *Biology and Philosophy*, *14*(3), 349–374.

Godfrey-Smith, P. (1999). Genes and codes: Lessons from the philosophy of mind?. In V. G. Hardcastle (Ed.), *Biology meets psychology: Constraints, conjectures, connections* (pp. 305–331). Cambridge, MA: MIT Press.

Griffiths, P. E. (1999). Squaring the circle: Natural kinds with historical essences. In R. A. Wilson (Ed.), *Species: New interdisciplinary essays* (pp. 208–228). Cambridge, MA: MIT Press.

Griffiths, P. E. (2001). Genetic information: A metaphor in search of a theory. *Philosophy of Science*, *68*(3), 394–412.

Griffiths, P. E. (2002). Lost: One gene concept, reward to finder. Essay review of *The concept of the gene in development and evolution*. In P. Beurton, R. Falk, & H. J. Rheinberger (Eds.), *Biology and Philosophy*, 17(2), 271–283

Griffiths, P. E. (forthcoming). Emotions as natural kinds and normative kinds. *Philosophy of Science, 71*(Suppl., *Proceedings of the 2002 Biennial Meeting of the Philosophy of Science Association, Pt. II. Symposia Papers*).

Griffiths, P.E. (forthcoming). The fearless vampire conservator: Philip Kitcher, genetic determinism and the informational group. In C. Rehmann-Sutter, & E. M. Neumann-Held (Ed.), *Genes in development: Rethinking the molecular paradigm*. Durham, NC: Duke University Press.

Griffiths, P. E., & Knight, R. D. (1998). What is the developmentalist challenge? *Philosophy of Science*, *65*(2), 253–258.

Griffiths, P. E., & Neumann-Held, E. M. (1999). The many faces of the gene. *BioScience*, *49*(8), 656–662.

Hacking, I. (1991). A tradition of natural kinds. *Philosophical Studies*, *61*, 109–126.

Medin, D. L. (1989). Concepts and conceptual structure. *American Psychologist*, *44*(12), 1469–1481.

Moss, L. (2001). Deconstructing the gene and reconstructing molecular develomental systems. In S. Oyama, P. E. Griffiths, & R. D. Gray (Eds.), *Cycles of contingency: Developmental systems and evolution* (pp. 85–97). Cambridge, MA: MIT Press.

Moss, L. (2002). *What genes can't do*. Cambridge, MA: MIT Press.

Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, *58*, 168–184.

Oyama, S., Griffiths, P. E., & Gray, R. D. (2001). Introduction: What is developmental systems theory? In S. Oyama, P. E. Griffiths, & R. D. Gray (Eds.), *Cycle of contingency: Developmental systems and evolution* (pp. 1–11). Cambridge, MA: MIT Press.

Rheinberger, H.-J. (1997). *Towards a history of epistemic things: Synthesising proteins in the test tube*. Stanford, CA: Stanford University Press.

Rheinberger, H.-J. (2000). Gene concepts: Fragments from the perspective of molecular biology. In P. J. Beurton, R. Falk, & H.-J. Rheinberger (Eds.), *The concept of the gene in development and evolution* (pp. 219–239). Cambridge: Cambridge University Press.

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* (pp. 187–232). Boston Studies in the Philosophy of Science, Vol. 183. Dordrecht: Kluwer Academic.

Schaffner, K. F. (1998). Genes, behavior and developmental emergentism: One process, indivisible?. *Philosophy of Science*, *65*(2), 209–252.

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

Waters, C. K. (1994). Genes made molecular. *Philosophy of Science*, *61*, 163–185.

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.