GENETIC VARIATION: DIFFERENCE, DEVIATION, OR DEVIANCE?

by

Lisa Anne Gannett
Department of Philosophy

Submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
April 1998

ABSTRACT

The dissertation focuses on scientific understandings of genetic variation in view of the Human Genome Project's (HGP) aim to map the estimated 50,000 to 100,000 genes and to sequence the approximately three billion nucleotides of the haploid human nuclear genome by the year 2005. There is legitimate concern that the "presumably representative" composite DNA reference sequence that is produced may institute a standard of genetic normality that treats departures from the sequence as at least potentially pathological and fails to appreciate the prevalence and propriety of genetic variation. Consideration of how the human mitochondrial DNA reference sequence has been used in different areas of biomedical research since it was published in 1981 reveals that it operates both as a statistical and a functional norm. I explore the evolutionary and clinical contexts that surround how genetic mutation, genetic variation, and genetic normality are understood in human molecular genetics. Evolutionary biologists and philosophers of biology have criticized the HGP for being anti-evolutionary in its treatment of genetic variation as deviation from a norm rather than simply as difference. I argue that these criticisms are mistaken in that the classical and neutralist theories of population structure authored by H. J. Muller and Motoo Kimura respectively present similarly normative treatments of genetic variation. From the clinical perspective, the question is whether genetic variation constitutes deviation from an objective biological norm or culturally constructed deviance. I argue that Georges Canguilhem's two-part thesis that knowledge of the pathological is antecedent to and constitutive of knowledge of the normal and that clinical judgements of health and disease precede theoretical judgements of biological normality and abnormality can be extended from physiology to human molecular genetics. Departing from Canguilhem, I conclude that judgements of genetic normality and abnormality, like judgements of health and disease, incorporate aesthetic, moral, social, and cultural, as well as biological, norms. Genetic explanations are influenced by extrascientific values in the additional way that they involve a pragmatic privileging of genetic over non-genetic factors that reflects social, economic, and clinical, as well as scientific, aims.

Keywords

genetic variation, genetic normality, genetic mutation, Human Genome Project, eugenics, reference sequence, wild-type, consensus sequence, classical-balance debate, neutralist-selectionist debate, Muller, Dobzhansky, Kimura, typological-population distinction, Canguilhem, health, disease, genetic explanation, genetic trait, pragmatic

ACKNOWLEDGEMENTS

It would have been infinitely more difficult to complete this dissertation and to survive graduate school without the intellectual and emotional support of three very special people. My mother, Grace Brundage Gannett, and my father, Jim Gannett, were formative influences on my development of interests in science, epistemology, politics, and social justice. They encouraged me in this project, as they have in all endeavours that I have ever taken on. Much of the thesis was written during months I spent at home in Toronto, a task greatly aided by good food and much kind indulgence. My advisor, Kathleen Okruhlik, maintained faith in my abilities at times when I had certainly lost it. Her "Feminist Epistemology and Critiques of Science" course in Fall 1991 led to my interest in pursuing a career in philosophy, specifically in the philosophy of biology. This was only in part because of the content of the course. Mostly, I was inspired by her unique combination of so many admirable traits: intelligence, integrity, empathy, insight, passion, warmth, and incisive critique. I would also like to thank Kathleen for sharing her office with me this past year and, along with Jim, Elizabeth, and Stephen, the basement suite of their home.

With Kathleen on my advisory committee were Paul Handford, Robert DiSalle, and, initially, Sahotra Sarkar. Paul's course in evolutionary genetics stimulated my interest in the classical-balance and neutralist-selectionist debates and, hence, Chapter Three of the dissertation. I am grateful for the detailed and thoughtful comments that I received from Paul on all chapters. My appreciation for a combined historical and conceptual approach to the philosophy of science was developed in an undergraduate course that I took from Robert in Spring 1992. I am thankful for Robert's insight and encouragement over these past years and for his comments on the prospectus and Chapter Five. Sahotra helped me a great deal with the prospectus, contributed comments on Chapter Five, and very considerately loaned me books and articles. Finally, although John Nicholas was not an "official" member of my advisory committee, he was a kind and more-than-capable "stand-in." I am also grateful for "Nick's" assistance in his capacity as graduate chair.

Like others in the department, I have been assisted a great deal by extremely capable and dedicated "front hall" administrative staff: Pauline Campbell, Melinda Ellison, Glenda Hunt, Julia McDonald, Andrea Purvis, Marie Schell, and Ella Young. I would especially like to thank Andrea for her assistance in sending out postdoctoral and job applications and for all of her support surrounding the stress that this involves.

My siblings — Lynn, Susan, and Dave Gannett — have been encouraging of my return to university to study (horrors!) philosophy. I have shared ups and downs with many friends, all of whom I cannot name here. Tara Abraham, Janet Childerhose, Jill Lazenby, Ken Meiklejohn, and Trevor Scanlan are the best "brunch buddies" I can imagine. I was incredibly lucky to enter the graduate program with a "cohort" the likes of Francisco Flores and Andrew Reynolds. Darcy Cutler was a great office-mate. I am grateful for the bed and late night chats at the home of Andrew, Kellie White, and cat Leonard. I shared much with Karen Detlefsen, Val Greifenberger, Moira Howes, Marcia Richards, and Christine Tetrault. I can always count on pre-philosophy friends like Karen and Charles Rice, as well as Tom Overend and the no-longer-kids we used to coach, to prevent me from taking myself too seriously.

During my graduate school career, I received financial support, for which I am very grateful, from the Social Sciences and Humanities Research Council (SSHRC) of Canada, the Ontario Ministry of Training and Education, and the University of Western Ontario. I also benefitted greatly by attending, on two occasions, the History of Biology Seminar at the Marine Biological Laboratory in Woods Hole. I would like to thank the Dibner Institute for the funding that I received, organizers Gar Allen, John Beatty, and Jane Maienschein, and my fellow participants with whom I continue to enjoy interacting.

DEDICATION

This work in dedicated to my parents, Grace Brundage Gannett and Jim Gannett, from whom I learned of the creative tension that exists between the humanities and the sciences.

TABLE OF CONTENTS

Chapter One

The Human Genome Project and Twentieth-Century Eugenics: The Meaning(s) of Genetic Variation

The century that opened with rediscoveries of Gregor Mendel's studies on the patterns of inheritance in peas is closing with a research project in molecular genetics that promises to be the initial, and necessary, step in attaining a complete understanding of the hereditary nature of humankind. The Human Genome Project (HGP) is a multi-billion dollar undertaking that aims to map the some 50,000-100,000 genes[1] and to sequence the approximately three billion nucleotide bases of the haploid human genome by the year 2005. The HGP is both basic science project and raw technological feat. Justifications for the project range from the esoteric to the practical. At the esoteric end of the scale are outcomes such as the development of a wholly theoretical biology (Gilbert 1992, p. 92), the self-understanding that will come with locating "those genes that make us uniquely human" (ibid., p. 94), and even "the total understanding of life itself" (Watson 1993, p. 312). Practically speaking, foreseen benefits include: improved understandings of human diseases, a twenty-first century "rational" "DNA (dioxyribonucleic acid)-based medicine" that tailors treatment regimens to individuals according to their genetic makeups (Caskey 1992; Hood 1992), and the ability to insert new favourable genes into the human germ-line. These optimistic outlooks are not shared by many who are concerned about ethical, social, and political issues surrounding the project. For example, Dorothy Nelkin and Laurence Tancredi (1989) and Ruth Hubbard and Elijah Wald (1993) point to other likely outcomes of the HGP: genetic discrimination in education, employment, and insurance; the resurgence of eugenics; support for genetic determinism;

[1] The total number of human genes remains a matter of much dispute. Taking a narrow definition of 'gene' as a segment of (not necessarily continuous) DNA that codes for a protein, recent estimates by leading molecular biologists include: 60,000-70,000; no more than 60,000; 80,000-100,000; "a lot more than" 100,000; and 120,000-150,000 (Cohen 1997, p. 769). In one scientist's words: "any number anyone gives you is just a wild guess" (ibid.).

the individualization of health and disease; and disputes over patenting and profiteering by biotechnology companies.

It is my aim in this dissertation to contribute to these critical discussions of ethical, social, and political issues surrounding the HGP and human molecular genetics research more generally. Specifically, I am interested in the potential for the HGP's penultimate product, "the complete sequence of a presumably representative human genome" (Maddox 1991, p. 11), to serve as a normative standard that treats intraspecific genetic variation as abnormal and unwelcome deviation and encourages and facilitates the development of biotechnological interventions that restore the norm. It is likely, though, that in all of human history no two individuals who are not monozygotic siblings have ever had identical DNA sequences. Can meaningful use therefore be made of a single DNA sequence as a reference for comparison with other sequences in view of the vast number of differences that will be encountered? Does finding sequence differences of likely functional significance constitute a directive to action that favours biotechnological intervention at the level of the genome? These questions indicate the two main components of the dissertation. First, in Chapters Two, Three, and Four, I analyze the concepts of genetic variation, genetic normality, and genetic mutation, focusing especially on the distinctions between normal and abnormal genetic variation and normal and mutant genes. Second, in Chapter Five, I turn to a phenomenon epidemiologist Abby Lipmann (1991) refers to as "geneticization" — the increasing tendency to understand human variation in terms of genetic variation. I approach the question of geneticization from the directions of genetic causation and genetic explanation: how is it that genes — whether normal or abnormal — can be said to cause or to explain traits? Chapter One provides a brief historical introduction to the HGP and considers contemporary research in human molecular genetics and the development of the new genetic technologies within the context of twentieth-century eugenics. The dissertation's concluding remarks examine the theoretical content contained in the concept of a "normal" genome and address the question whether such "normal" genomes actually exist, as well as the significance of any possible gap between idea and reality.

1.1 The Human Genome Project: Historical Background

The aim of the HGP is "to construct common resources for the study of human genetics" (Watson and Cook-Deegan 1990, p. 3322) by mapping and sequencing the entire human genome. The project will "create an encyclopedia of the human genome — a complete map and sequence" (McKusick 1989, p. 913) to serve as "a comprehensive source book for biology and medicine" (ibid., p. 914). The project has been conceived as a three-part plan: first, the creation of genetic maps; second, the production of physical maps; and third, the determination of the complete DNA sequence. The genetic, physical, and sequence maps are tools to be used to produce an additional map — the human gene map — that, at least at the beginning, was not explicitly part of the HGP. This comprehensive catalogue of the sequence and location of all human genes is expected to be ready at the same time as sequencing is completed. It is predicted that the molecular genetic knowledge that will arise from researchers' abilities to access map and sequence data will shed light on human disease: "There are over 4000 known human disorders inherited in a Mendelian fashion, and the outcomes of virtually all human degenerative and infectious diseases are influenced by the genetic make-up of the individual" (Gottesman and Collins 1994, p. 591). Another practical spinoff foreseen for the HGP is its consolidation of the United States' competitive edge in the international biotechnology industry — this aspect helped to convince the U.S. Congress to support the project (Macer 1991, p. 187).

Three scientists — Robert L. Sinsheimer, Renato Dulbecco, and Charles DeLisi — are credited with the idea of initiating a project to sequence the entire human genome. In May 1985, Sinsheimer, who was then chancellor at the University of California at Santa Cruz (UCSC), met with an expert group of scientists with backgrounds in DNA mapping and sequencing about the feasibility of setting up such a project at Santa Cruz, an idea he had begun to entertain the previous year. The impetus was Sinsheimer's desire not to see the university be forced to return a $36 million private donation.[2] The donation had been made to support the building of a telescope but the remaining necessary monies for the telescope were unavailable if it was to be named for the original

[2] All cost estimates in the dissertation are in United States dollars.

donor (see Cook-Deegan 1994, chapter five). It was Dulbecco who brought the idea of sequencing the human genome to a wide audience of scientists with his March 1986 commentary in *Science* where he urged a national sequencing effort of "comparable significance" and "carried out with the same spirit" as "the effort that led to the conquest of space" (p. 1056). Dulbecco had already presented the idea at two talks in 1985: in September at Cold Spring Harbor and in October at an Italian-American meeting in Washington, D.C. (Dulbecco 1993, p. 259). He believed that sequencing offered the best approach to studying cancer, as well as other pathological conditions and problems of physiology, development, and the nervous system generally, because it would make available any DNA probe that might prove useful to mapping genes and to identifying cells in which genes are expressed.

Around the time that Sinsheimer was contemplating a human genome sequencing initiative for UCSC, scientists associated with the U.S. Department of Energy (DOE) began to talk about doing the same. The DOE's interest in genome sequencing reflects its long term research into radiation-induced heritable genetic damage sustained by those who survived the atomic bombs dropped on Hiroshima and Nagasaki by the U.S. at the close of the Second World War. At a December 1984 DOE-sponsored meeting in Alta, Utah, the idea of sequencing the human genome arose in discussions about the difficulties researchers face in detecting heritable and inherited mutations in atomic bomb survivors and their children, as well as in other DOE programs charged with monitoring genetic damage due to low-level exposure to radiation and other environmental hazards (Cantor 1990, p. 49). Having a DNA reference sequence would make it possible to detect mutations directly — at the level of the genome. Subsequent to this, the October 1985 preliminary draft of a congressional Office of Technology Assessment (OTA) report concerning technologies to measure heritable mutations in humans, prepared by a project director who had been present at the Alta meeting, came across the desk of the newly appointed director of the DOE's Office of Health and Environmental Research, Charles DeLisi. DeLisi thought that it might be feasible for the DOE to take on a project to sequence the entire human genome. By the end of the calendar year, he had drafted a proposal for a Human Genome Initiative. A workshop was convened at Sante Fe in March 1986 for scientific discussion on the proposal. In May, DeLisi issued a funding

request to the DOE for a two-phase Human Genome Initiative that would begin by producing physical maps for each chromosome while working on the development of high-speed automated sequencing technologies and more advanced methods of computer analysis before proceeding to large-scale sequencing. The DOE genome program commenced in 1987 with reprogrammed internal funds totalling $4.5 million; however, continued funding was contingent on obtaining the support of the Senate and Congress (see Cook-Deegan 1994, chapter seven).

Some prominent molecular biologists like Walter Gilbert and James D. Watson supported a genome initiative from the beginning — Gilbert had been sold on the idea at Sinsheimer's May 1985 meeting and had subsequently taken on the task of garnering support for the project from other molecular biologists and the public. At a June 1986 conference at Cold Spring Harbor titled "The Molecular Biology of *Homo sapiens*," sequencing the human genome was a matter of much discussion. At an informal session held on the topic, some biologists supported the idea of a project dedicated to mapping and sequencing the human genome but expressed concerns about the DOE, rather than the National Institutes of Health (NIH), being at the helm. Others opposed the idea altogether, fearing a move to "Big Science" that would see the diversion of research funds from traditional single researcher-led projects to a small number of large laboratories geared to large-scale mapping and sequencing (Watson 1990, p. 45). In subsequent months, however, controversy over "[w]hether to start a genome project gave way to what it encompassed, how best to do it, and who should lead it" (Cook-Deegan 1994, p. 125). This transition was facilitated by a public forum hosted by the Howard Hughes Medical Institute in July 1986 and a scientific review of the proposed genome project carried out by a panel appointed by the National Research Council (NRC) of the National Academy of Sciences that began to meet in December 1986 (see Cook-Deegan 1994, chapters nine and ten). The NRC committee was a mix of those who supported and those who opposed the project but, in the end, there was unanimous agreement. The NRC report, issued in February 1988, recommended a fifteen-year program to map and to sequence the genome. The total price tag estimated for the project at completion was three billion dollars. Genetic and physical maps would be completed first with large-scale sequencing beginning in earnest only when the development of new sequencing technologies had

lowered sequencing costs substantially. The project would expand to include the mapping and sequencing of the genomes of several nonhuman "model" organisms. International cooperation was also emphasized. These modifications placated many of the molecular biologists who were initially opposed to the project.

It also helped that the NIH began planning its own genome program in the fall of 1986. Discouraged by the slow pace of progress, Gilbert attempted in 1987 to found a private company, Genome Corporation, dedicated to mapping and sequencing the human genome. However, uncertainties in the financial market at the time prevented the corporation from getting beyond the planning stages (see Cook-Deegan 1994, chapter six). Public monies were forthcoming: in 1988, the NIH received $17.2 million and the DOE received $12 million. From 1986 to 1988, the NIH and the DOE engaged in a leadership tug-of-war. Although the 1988 NRC report did not specify which government agency should take charge of the overall project, an OTA report commissioned by the House Committee on Energy and Commerce in 1986 and released in April 1988 recommended that an inter-agency task force coordinate the efforts of the two separate genome programs (Cook-Deegan 1994, p. 153). However, if the programs were to be united into a single project with only one agency in charge, the report recommended that it be the NIH (ibid., p. 160). Faced with the prospect of legislation to force their cooperation, a memorandum of understanding was signed by the two agencies in the fall of 1988 and a joint NIH-DOE advisory group was appointed (ibid., p. 167). At about this time, James Wyngaarden, director of the NIH, announced that Watson would serve as the first director of the Office of Human Genome Research that he had created at the NIH earlier that year.

From 1988 to 1990, the NIH and the DOE had genome research programs that operated independently of one another. The NIH received the lion's share of the funding: in 1989, $28.2 million went to the NIH and $18 million to the DOE; in 1990, these amounts increased to $59.5 million and $26 million respectively. Accompanying this increase was a change in status for the NIH's Office of Human Genome Research. As a "Center" — the National Center for Human Genome Research (NCHGR) — it had the authority to administer its own research grants. A joint five-year plan was released by the NIH-DOE subcommittee in April 1990. In conformity to the NRC report, priority in the first five years was to be given to the creation of genetic and physical maps with

large-scale sequencing waiting until sequencing efficiency was improved and the cost per base lowered. The plan established specific goals to be accomplished by 1995. Genetic linkage maps were to be completed with a resolution of 2 to 5 centimorgans (cM). Physical maps were to be completed with sequence-tagged site (STS) markers spaced approximately 100 kilobases (kb) apart and 2-megabase (Mb) contiguous overlapping clones ("contigs") assembled for large sections of the genome. Sequencing costs were to be reduced to $0.50 per base and ten million bases of contiguous DNA (0.3 percent of the genome) were to be sequenced. Watson announced the "official" Human Genome Project start date to be October 1990, the beginning of the 1991 fiscal year (Cook-Deegan 1994, p. 168).

Although the HGP's inceptions were in the U.S., it did not take long for it to become an international venture. Many European countries sponsor genome programs. Italy's genome program began as a pilot project in 1987 under the leadership of Dulbecco (Dulbecco 1993, p. 259; Cook-Deegan 1994, p. 187). Russia's genome program began in 1988 in the old U.S.S.R. (Cook-Deegan 1994, pp. 194-195). France began to fund genome research in 1988 and had developed a more centralized, although not very well-funded, program by 1990 (ibid., pp 195-196). Private sector initiatives in France have been more significant. Daniel Cohen and Jean Dausset founded the Centre d'Etudes du Polymorphisme Humain (CEPH) in 1983. CEPH subsequently undertook to coordinate a combined effort by several international research groups to construct a complete genetic linkage map of the human genome (ibid., pp. 43, 197). In 1991, Cohen and the French muscular dystrophy association (AFM) launched Généthon as an industrial-sized mapping and sequencing operation where a group led by Jean Weissenbach set out to create genetic markers for the entire genome and a group led by Cohen took on the task of compiling physical maps of all the chromosomes (ibid., pp. 196-197). Great Britain's genome program received its official start in 1989 although Sydney Brenner had commenced genome research at the Medical Research Council (MRC) laboratory several years before this. The British genome program was funded at the outset with public funds from the MRC and private monies from the Imperial Cancer Research Fund and, later, the Wellcome Trust (ibid., pp. 188-189, 211). Germany, haunted by its Nazi past, lagged behind other European countries. Although individual researchers received government

funds for genome research in the late-1980s and participated in the European Community (E.C.) initiative, no actual national genome program was undertaken until 1995 (Kahn 1996, p. 570).

The E.C. began to coordinate multinational efforts to map and to sequence the genomes of several "model" organisms in 1988 and tabled a research proposal for the human genome that same year. The proposal was modified following recommendations from Denmark and Germany and was adopted in 1990 (Rix 1991). The major genome player outside the U.S. and Europe is Japan. Led by Akiyoshi Wada, Japan began to fund the development of automated DNA sequencing technologies in the early 1980s. This five-year lead time over the U.S. in the research and development of sequencing technologies was instrumental in encouraging members of the U.S. Congress, concerned about U.S. global competitiveness in biotechnology, to support the NIH/DOE genome projects. The U.S. has since been critical of the Japanese government for inadequately funding genome research — especially basic scientific, as opposed to technological, research. At one point, in 1989, Watson threatened to withhold sequence data from the Japanese if they failed to increase their efforts (see Cook-Deegan 1994, chapter 15). Canada began a four-year genome program in 1992 with funding by government grant agencies and the National Cancer Institute (ibid., pp. 204-205). Because of federal budget cuts, funding has not been renewed (Kaiser 1997, p. 303). In September 1988, scientists formed an international body — Human Genome Organization (HUGO) — for the coordination of mapping and sequencing efforts in these (and other) countries with the goals of facilitating the transfer of information, assisting with international workshops, providing a forum for the discussion of ethical, social, commercial, and legal issues, etc. HUGO, described as a "U.N. for the human genome" (Zinder in McKusick 1989, p. 913), began to receive funds in 1990 from two private foundations: the Wellcome Trust in the U.K. and the Howard Hughes Medical Institute in the U.S. (Cook-Deegan 1994, p. 209). In 1990, UNESCO began to contribute funds toward fostering international cooperation in genome research that would include less wealthy nations from Eastern Europe and the "Third World" (Cook-Deegan 1994, p. 206).

Molecular biology's current focus on gene mapping and DNA sequencing reflects the importance that twentieth-century biology attaches to the gene's influence on

organismal development and behaviour as well as the physical reductionism that has shaped biology since the mid-nineteenth century. Suppositions that the gene's nature is chemical and that cellular activities and whole organisms are under genetic control date at least to the first quarter of this century (see Muller 1922). Until the early 1950s, molecular biologists believed that proteins were the only molecules of sufficient complexity to account for the properties of genes. Several experiments, beginning with Oswald T. Avery's work on the transforming principle in the early 1940s, showed DNA to be the more likely candidate (see Avery et al. 1944; Chargaff 1950; Hershey and Chase 1952). With the 1953 discovery of the double-helical structure of DNA by James D. Watson and Francis Crick and, eventually, success in "cracking" the genetic "code" in the 1960s (the complete genetic code was finalized in 1966), the attention of molecular biologists had long been fastened on the importance of nuclear DNA. But if the "conquest" of the genome was an idea whose time had come for molecular biology in the mid-1980s, this was only because of the technological developments that had occurred in the field from the early 1970s and on. As Crick remarked, reflecting back on the exciting early days of molecular biology in an address at a conference celebrating the fortieth anniversary of the discovery he shared with Watson: "if I had been asked if it would ever be possible to sequence the entire human genome, I would have predicted that this would take at least another hundred years" (1993, p. 18).

An impressive collection of tools began to be amassed by molecular biologists around 1970. Two different techniques for DNA sequencing were developed in the mid-1970s. Frederick Sanger, already a Nobel prize winner in 1958 for the protein sequencing method he developed in the 1940s and used in his nearly decade long project to determine the structure of bovine insulin, shared a second Nobel prize in 1980 for DNA sequencing. Sanger's method was first published in a 1975 paper co-authored with Alan Coulson. DNA polymerase is used to initiate complementary base pairing in solutions containing single-stranded DNA and free nucleotides. Four different reactions are set up: in each, one of the four nucleotide bases is missing and replication is incomplete. Since the base that would have been added next in the chain is known, the nucleotide base at each position in the sequence can be identified when gel electrophoresis is used to separate the fragments by length (Cook-Deegan 1994; Judson 1992). Sanger's modified "chain-

terminator" method, published in 1977, rather than "starving" the replication reaction, uses radioactively labelled dideoxy nucleotides which substitute for each of the four nucleotide bases during DNA replication and cause replication to cease (ibid.). Around the same time, Allan M. Maxam and Walter Gilbert introduced an alternate DNA sequencing method. Their method involves the use of a controlled chemical reaction which is capable of directly fracturing DNA at the sites of specific nucleotide bases. By comparing the length of fragments, as in Sanger's method, the nucleotide at each position in the sequence is identified (Cook-Deegan 1994). Pulsed-field gel electrophoresis, introduced by David Schwartz and Charles Cantor in 1984, improved the feasibility of large-scale sequencing because it permits large segments of DNA (up to ten million bases long as opposed to segments of up to thirty thousand bases with standard gel electrophoresis) to be sorted by length (Judson 1992, p. 74). In 1986, Leroy Hood and associates at Caltech and Applied Biosystems modified and automated Sanger's method. The automated DNA fluorescence sequencer labels each type of dideoxy nucleotide with a fluorescent, rather than radioactive, label. As the fragments are separated by electrophoresis, the fluorescent labels are excited by a laser and the information is stored in a computer (Cook-Deegan 1994, p. 66; Judson 1992, pp. 76-78).

Physical maps are necessary precursors to large-scale sequencing of the chromosomes. These maps order the DNA fragments to be sequenced by identifying unique physical markers (sequence-tagged sites or STSs) at regular intervals along each chromosome. Physical maps order DNA libraries, which are collections of DNA clones that permit DNA to be produced in the quantities necessary for sequencing. Several developments in the early 1970s made DNA cloning possible. In 1970, bacterial enzymes, called restriction enzymes, were discovered that cut DNA at specific sites (in nature, these enzymes protect bacteria from infiltrating viruses). When it was subsequently found that some restriction enzymes left fragments with "sticky ends" that would easily recombine and that the cell uses other enzymes for DNA repair, molecular biologists became able to cut and paste DNA and to combine DNA from different sources. Also, in the early 1970s, it was discovered that plasmids could be extracted from bacteria and then returned with an insert of foreign DNA. The bacteria would continue to multiply and replicate its own, as well as the foreign, DNA. In 1973, Herbert Boyer,

who had worked extensively with restriction enzymes, together with Stanley Cohen, who is credited with developing efficient methods for plasmid reinsertion, managed to use plasmids to carry animal genes into bacterial cells. As the bacteria multiplies, so do quantities of the gene (Judson 1996). With this, recombinant DNA technologies and the lucrative biotechnology industry were born. *In situ* DNA hybridization, invented in 1980, assists in physical mapping by allowing particular stretches of DNA to be located on the chromosomes. The relevant bit of DNA is produced in adequate amounts by cloning and a radioactive label is attached to fashion a DNA probe. Chromosomal position is revealed by observing where the probe hybridizes with separated strands of chromosomal DNA from genomic libraries (Judson 1992, p. 71). Early on, bacterial plasmid libraries were used for physical mapping. Yeast artificial chromosomes (YACs), in which DNA is attached to a much-reduced yeast chromosome and reintroduced into a yeast cell, were introduced in 1987 by Maynard Olson (ibid.) and bacterial artificial chromosomes (BACs) were developed in 1992 (Rowen et al. 1997). The advantages of YACs and BACs over bacterial plasmids is that much larger segments of DNA can be ordered and stored for sequencing.

Those who first conceived of a massive human genome mapping and sequencing project focused on constructing the physical maps that would make it possible to obtain the "ultimate" map: the complete DNA reference sequence. However, as planning proceeded, the importance of genetic linkage maps was increasingly emphasized because of their usefulness for gene mapping. Bacterial restriction enzymes have contributed to genetic, as well as physical, mapping. By 1980, researchers had discovered that the sites at which different restriction enzymes cut DNA are sufficiently variable among individuals that these restriction fragment length polymorphisms (RFLPs) could serve as DNA markers for the construction of human genetic linkage maps (Watson 1993, p. 310). Using RFLPs, the private Massachusetts-based Collaborative Research group led by Helen Donis-Keller, in a race with Raymond White's group at the University of Utah, published the first genetic linkage map of the entire human genome in 1987. The mapping and cloning of disease genes proceeds more readily as the density of markers placed on genetic maps increases and the HGP sought to improve the resolution of these early maps. It was also hoped that dense genetic maps would contribute to the identification of genes

involved in "complex" or non-Mendelian traits (Lander and Botstein 1986). In 1989, a new class of genetic markers, microsatellite repeats, was identified. Microsatellites are sets of tandem repeats of short (either dinucleotide, trinucleotide, or tetranucleotide) DNA sequences. Microsatellites quickly replaced RFLPs as the markers of choice for genetic linkage mapping because they are more highly polymorphic and are detectable by the polymerase chain reaction (PCR). In 1992, Jean Weissenbach's group at Généthon published a global genetic map that used microsatellite markers exclusively.

Several other technological developments have bolstered and redefined aspects of the genome mapping and sequencing initiatives. The polymerase chain reaction (PCR) was invented by Kary Mullis in 1985. With PCR, DNA can be multiplied *in vitro* instead of by cloning. The two strands of the DNA are separated by heating. Two bits of DNA are synthesized to be complementary to a specific short sequence at one end of the DNA sequence that is being amplified. These bind to the complementary sequences and serve as primers for polymerase enzymes to initiate DNA replication. As the reaction repeats, now beginning with two DNA molecules instead of one, an exponential amplification of the target sequence is initiated (Guyer and Koshland 1989). In only hours, a sequence of DNA can be amplified a millionfold and lots of material generated for sequencing. It was Marvin Carruthers in early 1980s who devised the method of synthesizing DNA strands of any desired base sequence, as is used to create primers for PCR. Carruthers' procedure was later automated by Leroy Hood. PCR has also been combined with reverse transcription to produce a powerful technique for mapping expressed genes. Reverse transcription was discovered independently in 1970 by David Baltimore, and Howard Temin and Satoshi Mizutani. These biologists discovered that transcription does not proceed only in one direction, from DNA to RNA, as per Francis Crick's Central Dogma. Some viruses use their RNA as a template to synthesize DNA. Reverse transcriptase, the enzyme that initiates this reaction, can be harnessed to produce DNA from mRNAs isolated from body tissues. The resulting complementary DNA (cDNA) differs from regular genomic DNA because, being complementary to mature mRNAs, it lacks introns that are transcribed into RNA but then edited out as well as regulatory regions that are not transcribed. Separate genomic and cDNA clone libraries are maintained. cDNA libraries are tissue-specific, for the liver, heart, kidney, etc.

The original NIH/DOE five-year plan was updated in 1993. The new five-year plan (in effect through 1998) accommodated progress that had been made in mapping, sequencing, and technological development since the first plan was formulated (Collins and Galas 1993). The original goal of a 2- to 5-cM genetic map was expected to be met by the 1995 target date. Indeed, Généthon's 1994 genetic linkage map, with more than 2000 microsatellite markers and an average spacing of 2.9 cM and only one gap larger than 20 cM, accomplished this one year early (editorial in *Nature* 1994). The 14 March 1996 publication of comprehensive genetic maps of "man and mouse" in *Nature* marked the end of the genetic mapping phase of the project. The 1996 human genetic linkage map has 5264 microsatellite markers located to 2335 positions with an average spacing of 0.7 cM (Jordan and Collins 1996, p. 111). Since the original goal of a physical map with STS markers at intervals of 100 kb would not be met by the 1995 target date — instead, an STS-based map with intervals averaging 300 kb was expected by 1995 or 1996 — the deadline was extended to 1998. In 1995, a preliminary global physical map was published as well as another physical map with 94 percent coverage from 15,000 markers (ibid., p. 112). Francis Collins, who took over as director at the NCHGR early in 1993 following Watson's April 1992 resignation, predicted in 1996 that physical maps would be completed in 1998 (ibid.). The 1993 five-year plan estimated that the projected goal of cost of sequencing of $0.50 per base might be met by 1996 but that the rate of sequencing would remain inadequate to meet the 2005 target date. The updated goal was to build up to a collective sequencing capacity of 50 Mb per year by the end of 1998 and to have 80 Mb of DNA (from both humans and "model" organisms) sequenced. This would be achieved by increasing the number of groups working on large-scale sequencing and heightening efforts to develop new sequencing technologies. By 1996, only one percent of the human genome had been sequenced. In 1995, the Wellcome Trust launched a $75 million seven-year concentrated sequencing project at the U.K.'s Sanger Center and, in 1996, the NCHGR awarded grants totalling $20 million per year for large-scale sequencing at a small number of laboratories in the U.S. (Marshall and Pennisi 1996). At the close of 1997, Collins proposed raising this contribution to $60 million per year (Wadman 1998). Other large-scale sequencing projects funded by governments or non-profit foundations are being carried out in France, Germany, and Japan. There are

also several corporate initiatives underway in the U.S. Although, in late 1997, only about two percent of the genome had been sequenced with the longest contiguous stretch of sequenced DNA in a public database at less than 1.5 million base pairs (Rowen et al. 1997, p. 605), the HGP's goal to sequence the entire human genome by 2005 is still believed to be attainable.

The 1993 five-year plan added a couple of new goals, both of which reflected technological changes as well as activities in the private sector. The identification of genes, and their incorporation onto physical and DNA sequence maps, became an explicit goal of the HGP. PCR with reverse transcription had been discovered to provide a rapid new method of gene identification. In the early 1990s, Craig Venter, at the time working for the NIH, had the idea of sequencing short regions of cDNAs — expressed sequence tags or ESTs — as a quick means of identifying and mapping individual genes. Watson's 1992 resignation from the NCHGR resulted from a conflict with NIH director Bernadette Healy concerning the NIH's application for patents on thousands of ESTs that Venter, who subsequently left the NIH for the private sector, had identified (Marx 1993; Thompson 1993). By 1993, serious efforts were underway in the private sector to partially sequence all cDNAs and to apply for patents on these ESTs. This had resulted in the withholding of private collections of ESTs from other researchers (Roberts 1993, p. 21). The inclusion of gene identification in the second five-year plan indicated the NIH's aim to compete with the private sector in gene mapping although many researchers believe that HGP funds should be confined to the provision of genetic and physical maps as the necessary infrastructure that allows others to pursue the genes (ibid., pp. 20-21).

A second new goal in the 1993 five-year plan was the development of technologies for the rapid genotyping that is necessary for medical research into complex non-Mendelian diseases and genetic "susceptibilities" to disease. This connects with a proposal made by Collins in the fall of 1997 that federal agencies, with possible private sector involvement, begin a systematic cataloguing of human sequence variation using single-nucleotide polymorphisms (SNPs). SNPs are common alterations in a single nucleotide in a stretch of DNA. They are better markers for genetic maps and for automated genetic scans than are microsatellites (Marshall 1997b). Genome variation could be catalogued, Collins suggests, by compiling SNP variants for individual genes

and/or by constructing a dense genome-wide SNP map to be used in identifying genes that contribute to complex traits. As he had for gene identification using ESTs, Collins expressed concern that private interests were collecting up SNPs and, with patents pending, withholding them from the public domain. The federal effort would place as many SNPs as possible in public databases where they can be accessed by researchers (Collins et al. 1997). Again, Collins' initiative has been criticized by some leading researchers who argue that the NHGRI[3] should concentrate HGP funds on sequencing the genome (Wadman 1998).

Collins' proposal to catalogue human genetic variation overlaps somewhat with the aim of the Human Genome Diversity Project (HGDP). This initiative was first proposed by Luigi Luca Cavalli-Sforza and others in a 1991 letter to *Genomics*. Their plan calls for DNA to be sampled from various isolated populations worldwide in order to be able to reconstruct human evolutionary history. The project targetted indigenous peoples and ethnic minorities and was proposed with some urgency because, as isolated populations increasingly merge with their neighbours, they begin to lose their distinct genetic identities (Cavalli-Sforza *et al.* 1991). HUGO took on responsibility for the HGDP early in 1994 (Knoppers *et al.* 1996, p. 272). But the project never got off the ground. The Ottawa-based Rural Advancement Foundation International (RAFI) alerted indigenous peoples' groups in early 1993 and the World Council of Indigenous Peoples (WCIP) unanimously denounced the project in December 1993 (Kahn 1994). Criticisms centre on the risks of commercial exploitation and "genetic colonization" implicated in the patenting of cell lines, the potential for genetic discrimination, the failure to appreciate non-western cultural values that view genes as sacred, and the lack of concern over the forseen extinction of these groups (Butler 1995). In 1995, UNESCO's International Bioethics Committee (IBC) failed to endorse the project (ibid.). The NIH and the NSF are currently prepared to fund only human genetic diversity research that originates in the U.S. until ethical, legal, and human rights issues are settled (Macilwain 1997b).

---

[3] In January 1997, the NCHGR became an "Institute" and was renamed the National Human Genome Research Institute (NHGRI) with yet additional control over research grants (Macilwain 1997a, p. 283).

1.2 Human Molecular Genetics and Twentieth-Century Eugenics

The acceleration of research in human molecular genetics means the identification of a constantly increasing number of genes and genetic markers associated with disease and dysfunction for which it will be possible to test individuals or screen populations. Since we all possess several genes that would be associated with serious diseases were they present in double rather than single dose, there is a huge potential market for carrier screening tests. Once couples "at-risk" for an affected offspring are identified, on the basis of their family histories or genetic tests, fetuses can be tested *in utero* and aborted if a genetic "defect" is found. Because knowledge of how genetic and environmental factors interact in the development of particular diseases lags far behind molecular genetics' successes in identifying disease-associated genes and genetic markers, prospective parents facing positive test results may perceive few alternatives to a decision to terminate the pregnancy. Social and economic factors may further constrain available choices. Alternately, embryos can be tested prior to implantation in *in vitro* fertilization (IVF). Where there is risk of hereditary disease, and especially if the individuals concerned (physicians and/or prospective parents) are opposed to abortion, IVF may be perceived to be an attractive option even if no infertility is involved. In the not-so-distant future, it may be possible to replace "defective" genes with "normal" genes or "normal" genes with "enhanced" genes in the early embryo or in the germ cells (ova and spermatozoa) prior to fertilization. Genetic manipulation of the germ-line in this way will affect not only the individual in whom the procedure is carried out but her or his future descendants. Of course, sometimes what counts as a "defective" genotype or a "defective" child rests in the eyes of beholders: quite apart from diseases that involve severe pain and/or early death, prospective parents might choose to abort a fetus or discard a preimplantation embryo that is likely to become a child who is insufficiently brilliant or of the wrong sex.

This impending scenario raises questions about the relationships of the HGP, and of molecular genetics research generally, with the history of eugenics. As we draw to the close of the twentieth century, what have the lessons of this century taught us? The early part of the century saw widespread public and scientific support for eugenic programs that

sought to control human breeding and the genetic characteristics of future generations. Contributions to future generations by such "undesirables" as the "feebleminded" and other mental "defectives," criminals, members of the lower classes, and members of racial and ethnic minorities were discouraged through such measures as education, immigration quotas, institutionalization, and involuntary sterilization. The "biologically fit" Anglo-Saxon middle and upper classes were encouraged through education, government financial incentives, and their senses of civic duty and entitlement to reproduce. Although support for eugenics in countries like England and the United States had waned by the eve of the Second World War, in part because of evident race and class biases, the tide of public opinion against eugenics fully turned once the horrors of the eugenic activities of the Third Reich became known. Physicians and leading scientists at German universities, propounding the theories of Mendelian genetics and evolution by natural selection, were complicit in Nazi measures that included the *Lebensborn*, mass sterilizations, and the exterminations of the handicapped, the mentally ill, homosexuals, Gypsies, and Jews.

It was inevitable that ethical discussions surrounding the HGP would at least in part be cast within the context of the past century of eugenics. Proponents of the HGP, as one might expect, emphasize the discontinuities between early and late twentieth-century eugenics. In announcing the creation of the ELSI program to study the ethical, legal, and social implications of the HGP out of NIH funds earmarked for genome research, Watson (1990) refers to "the terrible misuses of the incomplete knowledge of human genetics that went under the name of eugenics during the first part of this century" (p. 46) and the "vivid reminders" from Nazi Germany that "science in the wrong hands can do incalculable harm" (ibid.). Watson characterizes scientists who complied with the aims of the Third Reich as not just "bad guys" but bad geneticists, "servants of political and social masters" who practised pseudoscience (in Koenig 1997, p. 892). Historian Daniel J. Kevles (1995) similarly emphasizes the discontinuities between the eugenics of yesterday and of today. "Mainline eugenics" early this century was "flawed science" that incorporated race and class biases and inadequately understood the complexities of heredity. However, subsequent developments in human genetics, beginning with the "reform eugenics" movement of the 1930s and continuing after the war, have been favourable. Human genetics has been "emancipated" from previous race and class biases,

the science is "solid," its aims are medical not social, and interventions are justified in terms of the needs of individuals and individual families instead of their effects on such "abstractions" as the "race," "population," or "gene pool." Kevles believes not only that "scientific objectivity" has triumphed over "social prejudice" but that the contemporary social and political context with its emphasis on civil rights and civil liberties and its opposition to state-sponsored programs, as well as the existence of lobby groups representing members of minorities and those affected by disease and disability, will operate to prevent human molecular genetics from being turned to "eugenic ends."

Those who are critical of the HGP and wary of the biotechnological future are more likely to focus on the continuities of past and present efforts in human genetics and eugenics. Biologist and critic Ruth Hubbard, in her (1993) book written with Elijah Wald, points out that the eugenic belief that some people should have children and others should not persists; for example, surveyed physicians support sterilization more often when a woman is on welfare than when she is not, particularly if she has illegitimate children. Sociologist Troy Duster (1990) argues that although it appears progressive to have replaced studies of decreased intelligence in American blacks with those of increased genetic "susceptibilities" to multifactorial diseases such as lung cancer, heart disease, and mental illness, the effects are similar. Blaming genes draws society's attention away from unhealthy environments and weakens its commitment to address factors such as poverty, cigarette smoking (and tobacco advertisements), exposure to pollutants, and racism that contribute to these diseases. Hubbard (1990) reminds us that the Nazis tried out their gas chambers in hospitals before transferring them to the death camps, euthanizing the physically handicapped, the mentally ill, and homosexuals and then moving on to Gypsies and Jews. Phage geneticist Salvador Luria (1989) also does not shy from comparisons to the Nazis; Luria wonders if the HGP is not just a "kinder gentler program" than what the Nazis carried out — a program "to 'perfect' human individuals by 'correcting' their genomes in conformity, perhaps, to an ideal, 'white, Judeo-Christian, economically successful' genotype" (p. 873). Hubbard and Wald question the likelihood of genuine reproductive autonomy. First, the range of possible choices is constrained by the social supports that are available to persons with disabilities and their families. Second, many attending physicians will regard certain choices — the refusal of prenatal tests or the

refusal to abort following a positive test — as irrational or unacceptable. Third, the economic costs of caring for disabled individuals may mean that women find themselves pressured or mandated by their insurance companies or governments to undergo prenatal tests and abortion.

It is important to try to situate current developments in human molecular genetics within the context of the history of eugenics in the twentieth century. Scientists today have unprecedented control over hereditary material. The technological capacities to discern the genome's fine structure, to manipulate DNA, and to modify genes are powers that their predecessors could only dream about. And, despite movements away from "eugenics" after the Second World War, at no time during this century has the majority of geneticists veered from the conviction that genetic differences contribute significantly to individual differences in both mental and physical traits; that racial and ethnic groups differ in relative gene frequencies; and that human control over heredity and evolution is a desirable aim. But whether analysts conclude, with Watson and Kevles, that human molecular genetics is innocent of the taint of eugenics and contributes to the common good by fighting against disease, or, with Hubbard and Luria, that clinical genetics is just a "gentler and kinder" eugenics, there are problems in how exactly we are to understand the term 'eugenics.' As political scientist and historian of genetics Diane B. Paul so well points out: "'Eugenics' is a word with nasty connotations but an indeterminate meaning" (1994a, p. 143). Disagreements exist over whether to characterize eugenics according to a program's intentions or effects, its use of coercive rather than voluntary means, or its appeals to social and political aims that extend beyond the immediate concerns of individual families (Paul 1994a,b). In view of the difficulties of reaching a historically appropriate, sufficiently nuanced, and value-neutral definition of 'eugenics,' Paul (1994b) suggests that, if society is to grapple effectively with the eugenic implications of the HGP, the best approach may be to consider the likelihoods that certain scenarios people fear to be eugenic will arise.

I believe that Paul is right to question the usefulness of the label 'eugenics' — whether claimed or disavowed — in today's political debates over human molecular genetics and genetic medicine. Where history is used as "a weapon in a war over social policy" (Paul 1995, p. 134), it can indeed be a bludgeon. I believe that Paul is also

correct in emphasizing important discontinuities between applications of knowledge in human molecular genetics today and eugenics fifty years ago. Writing from a U.S. perspective, she concludes that eugenics is far less likely to be associated with government coercion, as it was early this century, than to result from the forces of a market economy — consumer demand for "better babies" in the case of positive eugenics and the economic interests of health management organizations (HMOs), insurers, and biotechnology companies, as well as physicians' litigation fears, in the case of negative eugenics (1994a, 1994b, 1995). Nevertheless, I will paint two broad strokes, in the next few pages and in the remaining dissertation, that highlight basic continuities in the relationships between genetics as theory, genetics as practice, and social values, over the course of this century. The first broad stroke targets assumptions that geneticists just discover "the facts" and that values enter only when scientific knowledge is applied in the practical domain. I argue that basic genetics research is not undertaken in the absence of practical aims and that extratheoretical values inform knowledge claims. This is true of human genetics and eugenics early this century; it remains true in human molecular genetics and clinical genetics today. The second broad stroke focuses on the judgements of the relative worth of different human lives that occur in formulating coercive governmental eugenic policies as well as in making individual reproductive decisions. Often such judgements are justified in terms of theoretical distinctions between "normals" and "mutants" and "healthy" and "defective" genes.

Watson and Kevles rigidly differentiate facts from values and science from its applications in distinguishing today's clinical genetics from yesterday's eugenics. They do so in several ways. One way is to treat eugenics as "flawed" or value-laden science and scientists complicit with Naziism as "bad" geneticists. Human genetics earlier this century was "bad" science insofar as it incorporated race and class biases that lent support to antimiscengenation and restrictive immigration laws in the U.S. and, most extremely, the "Final Solution" in Germany. Kevles considers human molecular genetics' focus on traits of clinical, rather than social, importance to demonstrate its "emancipation" from such biases. Another way is to regard past knowledge in human genetics to be "incomplete." Failures to recognize the complexity of gene-gene and gene-environment interaction and the "polygenic" basis of low intelligence were responsible for

"misinformed" eugenic practices like the segregation and forced sterilizations of the "feebleminded." Today's science is "solid" and the molecular approach promises a (soon to be?) complete knowledge of genetics: Gilbert (1992) envisions a wholly theoretical biology deduced from "axiomatic" DNA sequence data that predicts and explains all human development and behaviour. Yet another way is to warn of the dangers of letting "good" science fall into the "wrong" hands of those who will "misuse" it — we must maintain "vigilant" guard to prevent knowledge in human genetics from falling into the "wrong" hands ever again, says Watson. The HGP places knowledge in the "right" hands of molecular geneticists and clinicians who will use it to relieve suffering and to benefit humanity.

There are three sets of problems with this account. First, the historical reconstruction is itself "incomplete." As early as the 1910s, geneticists like T. H. Morgan and H. J. Muller had stressed the complex — many-one and one-many — relationships between genes and traits. R. A. Fisher understood very well, even in 1918, that intelligence is polygenic and yet still supported sterilizing the "feebleminded." That James F. Crow, writing in 1972, would urge genetic counsellors to discourage parents of low intelligence ("polygenic" in origin) who already have a child of low intelligence from reproducing shows that efforts to restrict reproduction in the "feebleminded" from the 1910s to the 1930s were not an aberration resulting from the inadequate knowledge of the time. Nor can Fisher or Crow, two extremely well-respected mathematical geneticists, be characterized as scientifically misguided! Second, that human molecular genetics focuses on clinical, rather than social, traits represents no guarantee that social values do not remain influential. A point of clarification is warranted at the outset. Although it is true that "single gene" diseases were initially found most tractable by the new molecular techniques, with more refined methodologies and increasingly powerful techniques, the molecular "dissection" of complex traits such as intelligence is possible. But, even so, it cannot be assumed that social values do not influence clinical judgements of disease and disability. Many traits considered previously to be moral or social are now regarded as medical — alcoholism and drug abuse, for example. Thus, Kevles' distinction between a social eugenics and a medical molecular genetics is as much a reflection of the territorial expansion of medicine as of geneticists' new social sensitivities. There is also

an implicit assumption that judgements of disease and disability are wholly based in a value-neutral biology. Insofar as diseases and disabilities represent departures from normal function, clinical interventions that seek to restore what is natural are justified. I argue in Chapter Four that moral, aesthetic, social, and cultural values are also implicated in judgements of normal and abnormal biological function and health and disease. Third, I am a great deal less confident than Watson that we so easily distinguish "right" from "wrong" hands. My major criticism, though, concerns the implied separation of different sets of pairs of hands — not those that would apply knowledge in human genetics "rightly" or "wrongly" but those that "do" science and those that "apply" science. This is the conceptual foundation that underlies the HGP's ELSI program. The program is dedicated to the ethical, legal, and social *implications* of the HGP and knowledge in human molecular genetics. The term 'implications' suggests that all significant questions of value arise consequent to the science and that values neither shape nor constrain science, nor are constituted in the doing of science. Are we really to believe that geneticists operate in a cultural vacuum in a way that is devoid of practical purpose?

Desires to understand heredity and desires to control heredity to fulfil certain ends have always been inseparable. The scientific study of heredity received its start in a practical setting driven by economic interests: the agricultural breeding of plants and animals. The scientific study of human heredity arose alongside and was itself directed by practical social and economic aims. Francis Galton is the founder of both human genetics and eugenics.[4] In 1883, Galton coined the term 'eugenics' which derives from the Greek word for "well-born." While the idea of a eugenic society goes back at least to Plato's *Republic*, and humans have no doubt exercised their preferences for offspring with some qualities rather than others by controlling marriages and selecting mates for even longer, with Galton it became scientific. Galton quantified traits, collected data, traced family pedigrees, developed statistical tools of analysis, and proposed the "law of

[4] Here I use 'human genetics' understood in a broad sense as the scientific study of human heredity. More narrowly construed, 'human genetics' represents the discipline that was founded in 1930 in the U.S. and in the U.K. to study the genetics of human diseases and behaviours. Although human genetics presented an alternative to a racist and classist eugenics, their memberships overlapped considerably (see Kevles 1995 and Paul 1995).

ancestral heredity." Since Galton, across vast changes in methodological approaches to the study of heredity, understanding and controlling heredity have remained intimately tied to one another. One needs only to recall Mendel and his peas to appreciate that the study of heredity has never been a strictly observational or theoretical science. It is difficult to identify prominent geneticists who have had no interests in the practical applications of their research, whether in agriculture or in human society.

An awareness of the potential for genetics to contribute eventually to human betterment has been in the back of the minds of many geneticists studying nonhuman organisms. Experimental organisms like peas, guinea pigs, *Drosophila*, *Neurospora*, *E. coli*, and bacteriophage that are considerably more tractable to study than humans have made it possible to uncover the basic laws and mechanisms that underlie hereditary transmission and gene action. Admittedly, there are biologists who are drawn to genetics solely by their desires to understand nature in and of itself. As well, some laboratory geneticists harbour great attachments to "their" experimental organisms with nary a thought to the significance of their research for humans. But there are many prominent examples of geneticists who have been motivated in their research with nonhuman organisms by the eventual importance the science of heredity would have for humans — for example, Theodosius Dobzhansky and H.J. Muller, both of whom spent their careers studying *Drosophila*, state this explicitly. According to Evelyn Fox Keller, after World War II, and the turning of the tide against eugenics,

> there was not talk about human genetics. All discussion of genetics was cast in the terms of basic science, and scientists were looking at organisms that are very far from human beings. You can't get much further than *E. coli*. Yet even in the early days of molecular biology it was clear that there was nothing distinctive about *E. coli*. They were studying *E. coli* as a model organism for all organisms. Monod's remark about "What's true of *E. coli* is true of the elephant," if it had been said at another time, would have been "What's true of *E. coli* is true of the human being."
>
> They were interested in genetics — we always have been interested in genetics — not just out of abstract interest in how the world works but very much out of self-interest in how *we* work. This is now explicit in the talk about the Genome Project. (interview with Casalino 1991, pp. 113-114)

With the HGP, a selection of experimental organisms have become "model" organisms. If early molecular biologists studied "*E. coli* as a model organism for all organisms,"

many molecular geneticists today study *E. coli*, one of the HGP's "model" organisms, as a model specifically for humans. Organisms serve as "models" for humans in several ways. Homologous regions in simpler organisms often help researchers to determine the functional significance of sequence data in humans. Human genes can be isolated and inserted in the genomes of "model" organisms to attempt to elucidate their functions and patterns of expression. Recombinant mice offer experimental models to study specific human diseases such as cancer.

The constant and ongoing relationships between geneticists' desires to understand heredity and their desires to control heredity, especially human heredity, leave me sceptical about another of Kevles' distinctions between medical genetics and eugenics. I agree that there is an important — conceptual as well as historical — distinction to be made between eugenicists who sacrifice individual well-being for the good of the whole by focusing on the "race," "population," or "gene pool" and medical geneticists who attend to the health of prospective individuals and the needs of families. But Kevles goes beyond this. He believes that medical genetics is to be embraced over eugenics because it embodies socially progressive tendencies. To understand the transfer of the locus of intervention from society to individual to be a transparent good, however, is just to prefer one political ideology to another. The liberalism that takes the rights and freedoms of individuals to be paramount coincides with, as Paul (1995) notes, the "remarkable transformation in public attitudes toward reproductive responsibility" that took place during the 1960s and 1970s (p. 129). Reproduction became private, a matter of individual rights to be protected from state intrusion, and genetic counsellors redirected their attentions from the long term effects of individual reproductive choices on the "gene pool" to the immediate desires of their clients for normal healthy children. Important technological developments in the areas of reproductive physiology and molecular and clinical genetics also occurred during these two decades: amniocentesis accompanied by prenatal genetic tests and the option of abortion; genetic engineering and cloning; recombinant DNA technologies; *in vitro* fertilization. These technologies made an entirely new locus of intervention in clinical genetics possible: the individual genome.

Over the course of the century, the site of intervention in eugenics/clinical genetics has shifted from population to family to individual in a way that parallels technical and

methodological changes in experimental genetics. Early twentieth-century eugenicists sought to intervene at the level of the population to restrict breeding within and between different "types" or "races" of humans. Classical geneticists, during this period, similarly sought to control breeding between mutant strains or "races" of *Drosophila* in order to discover basic mechanisms of hereditary transmission by tabulating the frequencies with which traits appeared in the progeny. In the 1950s, clinical geneticists began to be able to offer carrier screening tests for a limited number of conditions to prospective parents considered at-risk for an affected child due to their family history. Thus, the site of intervention moved from the population to the family. Screening tests that detect heterozygosity at a gene locus due to the presence of both normal and mutant forms of a protein in the blood of an apparently healthy individual became possible after biochemists began to develop the technical means to identify variations in protein structure in the late-1940s. Amniocentesis was developed in the 1960s and by the mid-1970s, with the availability of abortion, had become routinely used in prenatal screening. The site of clinical intervention became the individual: diseases could be "prevented" by the selective elimination of affected fetuses. Initially, prenatal screening involved biochemical tests and karyotyping; today, an ever-increasing number of genetic variants associated with disease can be tested for directly at the level of the genome. The availability of techniques such as chorionic villus sampling (CVS) and IVF increase the range of prenatal options available to women. Successes with recombinant DNA technologies over the past twenty-five years in the laboratory and in agriculture suggest that the genome as the preferred site for clinical intervention may not be far off.

In 1969, just as the technological revolution in molecular biology was beginning, Robert Sinsheimer, the molecular biologist who, as we have seen, was one of the originators of the idea to launch a massive human genome sequencing project, envisioned a "new eugenics" quite different from that of Galton. Sinsheimer conceived that the power of technology would make this "new eugenics" far superior to the old. Galton's eugenics relied for its success on the social control of breeding in successive generations and was limited in what it could accomplish. It could do no more than to improve the *relative frequencies* of *already existing* traits in the population. The "new eugenics," on the other hand, can be carried out in individuals all of whom it is possible to convert, at

least in theory, "to the highest technological level" (in Kevles 1995, p. 268). Wholly new genes and traits can be fashioned. By taking technological control of their future evolution, humans no longer have to settle for what nature has come up with. This is not a new vision. As early as 1916, *Drosophila* geneticist H. J. Muller foresaw that one day humans would be able to control evolution by learning how to control mutagenesis. Developing an analogy between biology and physics, Muller wrote: "Mutation and Transmutation — the two keystones of our rainbow bridges to power!" (in Keller 1990, p. 397). Keller (1990) argues that molecular biology's current "technological prowess" was not only anticipated in predecessors like Muller but is the consequence of "the forms of knowledge that biology, following physics, has taken as its norm" (p. 408). These "forms" include: "belief in the absolute adequacy not simply of materialism, but of a particular kind of (linear, causal) mechanism; belief in the incontrovertible value of simplicity; belief in the simultaneous equations between power and knowledge, and between virtue and power" (ibid., p. 407). A medical genetics that has abandoned "gene pools" for individual genomes is the not very surprising outcome of molecular biology's commitments to methodological and metaphysical reductionism. And if there is any truth at all in the thesis that eighteenth century liberal individualism and capitalism influenced modern science's adoption of a reductionistic metaphysics (see Lewontin 1993, pp. 10-12), it is not surprising that the practical applications of a molecular biology committed to understanding whole organisms only as the sum of their parts would remain consonant with these political and economic values.

I referred earlier to a second broad stroke I wish to paint. This concerns judgements that occur in applied areas of genetics — whether we call these eugenics, clinical genetics, or medical genetics — to distinguish between health and disease and favourable and unfavourable traits. As Paul notes, any new eugenics, positive or negative, is likely to be the result of market forces rather than government coercion. For instance, consumers may well demand such genetic services as carrier screening and prenatal genetic tests in order to ensure "normal," "healthy," or "better" babies. The desires of prospective parents for "normal" "healthy" babies are likely to be reinforced (if not forced) by physicians who wish to avoid malpractice suits and public or private insurance companies that prefer the cheaper cost of an abortion to the long-term support of a person

with a disability. Troy Duster (1990) refers to a "back door eugenics" and Robert Wright, writing in the *New Republic*, to a "homemade eugenics" (in Paul 1994b, p. 152) in their warnings that the parental decisions associated with the type of model that Kevles lauds — with its focus on the prevention and treatment of disease in individuals and private reproductive choices — remain eugenic since they are guided by aims to eliminate "defective" fetuses (Duster 1990, p. 128). Paul (1994a,b) argues that fears of a "back door" or "homemade" eugenics treat questions of eugenics in terms of consequences — the population-level effects of individual reproductive choices. But others believe that prenatal diagnosis means that eugenics is already with us since, as Philip Kitcher (1996) writes, "[e]ugenic practice begins with an intention to affect the kinds of people who will be born" (p. 193). Similarly, according to Abby Lipmann (1991): "prenatal diagnosis necessarily involves systematic and systemic selection of fetuses, most frequently on genetic grounds.... Prenatal diagnosis presupposes that certain fetal conditions are intrinsically not bearable" (pp. 24-25).

I am interested in the theories and values that inform, and provide justifications for, choices concerning what constitutes a "better" baby and what kinds of "fetal conditions are intrinsically not bearable." Although more momentous, these are similar in kind to decisions that lead us to seek medical advice, for ourselves and for our children. Theoretical medicine defines health in terms of normal biological function and disease in terms of abnormal biological function. The distinctions between health and disease and normal and abnormal function, that coincide also with the line between positive and negative eugenics, are often appealed to in order to provide ethical justification for genetic interventions. For example, prenatal genetic screening, whether of preimplantation embryos in IVF or of *in utero* fetuses where abortion is an option, is justified when the aim is to prevent the births of children with hereditary diseases. Similarly, germ-line manipulation is justified if the aim is to eliminate mutant genes associated with disease or dysfunction from a family or population. The legitimacy of the distinctions between normal and mutant genes, health and disease, and positive and negative eugenics is frequently taken for granted. For example, bioethicist Burke K. Zimmerman (1991) argues that "[t]he object of germ-line therapy should ... be to restore an 'original' healthy genetic topology to the treated individual, such that future

procreation would proceed as if one's progenitors had never carried a genetic lesion" (p. 599). Along the same lines, Alex Mauron and Jean-Marie Thévoz (1991) support germ-line intervention that has "*bona fide* therapeutic purpose" and "merely aims at restoring an order of things that obtained previously, but was disturbed by genetic mutation" (p. 656).

This uncritical acceptance by ethicists of concepts such as "'original' healthy genetic topology" and "genetic lesion" or "genetic mutation" is criticized by Camille Limoges (1994). Limoges argues that ethicists who fail to question the separation of scientific knowledge from its applications by placing themselves "downstream" of science ignore the ethical content of concepts like normality and mutation that arise "upstream" where they are "contrived and put to use" by scientists. If such concepts are to be subject to critical examination, philosophers concerned with ethical and social issues surrounding the new genetic technologies must move "upstream":

> it would seem that some issues of considerable social and ethical relevance are to be examined far *upstream* from where most ethicists intervene.... It ... underscores the limited effectiveness of a *downstream* bioethics conceived of as a rational discussion to help delineate a course of action regarding the suffering individual, or regarding the use of a technology. Key questionings occur, and ought to, *upstream*, while and where the science is being done. (p. 124)

Limoges is right to find bioethicists firmly ensconced on the "downstream" side of science on the value side of the fact-value divide. More often than not, bioethicists leave biology to the biologists, and even medicine to the medical doctors. Value theory provides the resources for rational decision-making most often concerning specific aspects of clinical practice and the applications of medical technologies — for example, regarding the new genetic technologies, questions about consent, abortion, access to genetic technologies, the potential for genetic discrimination, etc. But bioethicists are not the only group of philosophers vulnerable to Limoges' criticisms. Any approach that maintains rigid distinctions between facts and values, theory and practice, and science and technology, fails to engage in the "key 'upstream' questionings" that Limoges urges. If there is a tendency for bioethicists to remain "downstream," there is a reciprocal tendency for philosophers of science to situate themselves "upstream." Empirical facts and abstract scientific theories are their preferred company; they are quite happy to cede questions

surrounding the applications of science to ethicists or social theorists. However, although situated "upstream," insofar as philosophy of science adheres to traditionally rigid demarcations between facts and values and "upstream" science and "downstream" applications, many types of "key questionings" do not occur. The possibility that fundamental scientific concepts may incorporate normative, even culturally-laden, content is not entertained if the philosopher's accepted (and acceptable) project is to guarantee the objectivity of the science by securing its foundations in theoretical or empirical definitions of such concepts. Objectivist philosophers of medicine take a similar approach, seeking to furnish value-neutral foundations for medical practice in a theoretical medicine that is itself founded in theoretical definitions of the concepts of health and disease in terms of the empirical concepts of normal and abnormal biological functions. Clinical interventions are justified, in this sense, if they seek to eliminate disease and dysfunction and to restore nature's proper order.

My goal in the dissertation is to examine the distinctions between normal and abnormal genetic variation, "original" and "disturbed order," and normal or healthy and mutant or defective genes by engaging, as Limoges recommends, in "key questionings ... *upstream*, while and where the science is being done." This means adopting an approach that engages conceptual issues in human molecular genetics from both historical and philosophical perspectives and at the same time maintains sensitivity to ethical, social, and political contexts that may inform this analysis. Many traditional philosophical approaches, whether taken by bioethicists, philosophers of science, or philosophers of medicine, are inadequate to this task. My project recognizes that, in human molecular genetics, the concepts of normal and abnormal genetic variation and normal and mutant genes are of central theoretical importance *and* are also profoundly ethical insofar as they implicitly or explicitly justify genetic interventions in the practical realm. Consequently, for the purposes of the dissertation, I take conceptual analysis to be a critical project in which the boundaries that separate facts from values, theory from practice, and science from its applications are themselves at issue.

1.3 Genetic Variation: Difference, Deviation, or Deviance?

The title of the dissertation, *Genetic Variation: Difference, Deviation, or Deviance?*, alludes to three different possible valuations of genetic variation: value-neutral statistical difference; deviation from an adaptive biological norm (a "weakly normative" valuation); and deviance from nonbiological moral, aesthetic, social, or cultural norms (a "strongly normative" valuation). This is one way in which the dissertation engages the fact-value distinction. Genetic variation is sometimes regarded simply as difference, as neutral fact. Statistically normal genes may be considered to be those alleles that already exist in a population. Abnormal or mutant genes arise when the existing gene structure is in some way modified. Statistically normal genes may also be considered to be those alleles that are found frequently in a population. The normal gene need not be the *most* frequent allele at a given locus; rather, allelic variants at the same locus may be considered to be normal if they are frequent *enough*. Abnormal or mutant genes, then, are statistically rare alleles. Genetic variation may instead be considered as deviation from a biological (functional) norm. Functional norms can be understood in several different ways. The functional norm may be the "proper" function for which a gene was favoured by natural selection in the past or it may be a function of the gene that contributes positively to an organism's present fitness, that is, its relative ability to survive and to reproduce. Present organismal fitness may be conceived either in terms of average fitness (a typical genotype in the population) or superior fitness (the best genotype in the population). The functional norm may also be understood as a theoretical ideal that would optimize fitness given certain genetic and environmental parameters. Conversely, mutant or abnormal genes may fail to perform the function for which they were selected in the past, may have a fitness inferior to the average fitness of alleles at the locus, may have a fitness that is less than the best allele at the locus, or may have a fitness that is less than what would be ideal or optimal for an allele at the locus given a particular genotypic and environmental background. Lastly, understanding genetic variation as deviance rather than as deviation involves the recognition that the biological values of survival and reproduction are not the only norms that guide judgements of what counts as health or disease and normal or

mutant genes. Moral, aesthetic, social, or cultural values may also be implicated in such judgements.

Chapter Two, "Defining Genetic Normality and Denying Genetic Variation: The Human Genome Project's 'Presumably Representative' Human Genome," addresses the question of whether the HGP defines a standard of genetic normality that denies the presence, prevalence, or propriety of genetic variation. First, I consider the three types of maps that the HGP aims to construct — genetic maps, physical maps, and the sequence map. I argue that, for the most part, genetic and physical maps are tools for mapping genes and sequencing genomes and are not themselves standards of genetic normality that deny the presence, prevalence, or propriety of genetic variation. The sequence map, however, does have this potential. Second, I assess the likelihood that the completed nuclear DNA sequence map will represent a standard of genetic normality by investigating how the human mitochrondrial DNA (mtDNA) reference sequence has fared since it was produced in 1981. Third, I present two twentieth-century concepts of genetic normality — the concept of wild-type that dates to the early days of classical genetics and the concept of consensus sequence that is associated with contemporary human molecular genetics — in an effort to elucidate the nature of the normativity that is likely to attach to a human genome reference sequence. I emphasize the different senses of 'normality' expressed in the concepts of reference sequence, wild-type, and consensus sequence: statistical notions of normality with the normal conceived temporally as the already-existing or original value or spatially as the frequent, common, usual, or typical value; the normal or Gaussian distribution curve and the measures of central tendency it supports; functional notions of normality as "what works" and what contributes to organismal fitness (survival and reproduction); and the normal as the ideal. I place these functional and statistical notions of biological normality within the context of nineteenth-century developments that are associated with Claude Bernard, Adolphe Quetelet, and Francis Galton.

Conflations of these different senses of 'normality' in the concepts of reference sequence, wild-type, and consensus sequence would, in realms apart from biology, indicate serious semantic confusion. However, because biological entities have evolved by natural selection, it is understandable that statistical, functional, and even ideal notions

of normality would intersect. Chapter Three, "The Evolutionary Context: Is Genetic Variation Difference or Deviation?," concerns different evolutionary understandings of the concepts of genetic variation, genetic normality, and genetic mutation. I argue that human molecular genetics' typological treatment of genetic variation as deviation from a norm is not anti-evolutionary, as some critics have suggested, but rather is consistent with a *particular* set of evolutionary beliefs. To illustrate this, I focus on two key controversies in evolutionary theory regarding the genetic structure of populations: the classical-balance debate and its historical and conceptual successor, the neutralist-selectionist debate. The classical and neutralist positions are in agreement that genetic variation is either selectively neutral or harmful and very rarely beneficial, that genetic mutations are almost always deleterious, and that it makes sense to talk of a normal gene or genome because constant selective values can be assigned to individual alleles regardless of their genetic and environmental backgrounds. The balance and selectionist schools, on the other hand, agree that genetic variation is very seldom of neutral selective value, that genetic variation is beneficial to a population in both short and long terms, that constant selective values cannot be assigned to individual alleles because of the prevalence of gene-gene and gene-environment interactions, and that there is an array of normal genes and genomes. I conclude the chapter by arguing that a "bean bag" approach to population genetics cannot furnish evolutionary support for clinical applications of human molecular genetics because of the importance of gene-gene and gene-environment interactions in individuals.

Human molecular genetics tends to forget not only the evolutionary, but also the cultural, contingency of any standard of genetic normality. Chapter Four, "The Clinical Context: Is Genetic Variation Deviation or Deviance?," considers the relationships between human molecular genetics, clinical medicine, and culture. Functionalist theories of health and disease define these concepts objectively in terms of normal and abnormal biological functions. Clinical interventions are justified where the aim is to restore what is "natural." In this way, it is assumed that the normal is epistemically prior to the pathological and that theoretical knowledge in biology precedes practical action in medicine. In *The Normal and the Pathological*, Georges Canguilhem argues just the reverse: that practical action in the clinic precedes theoretical knowledge in physiology and that the pathological is epistemically prior to the normal. I argue that the situation

is similar in human molecular genetics. Without disease phenotypes, it would be impossible to identify normal gene structure and function. Consequently, to the extent that judgements of health and disease are value-laden and incorporate, besides biological values, moral, aesthetic, social, or cultural values, so too are judgements of genetic normality and genetic mutation. The scientific legitimacy of a disease designation cannot be verified by identifying a corresponding genetic mutation because what counts as normal or abnormal genetic variation and as a normal or mutant gene follows from an antecedent normative judgement about what counts as a disease. Disease judgements, I contend, are irreducibly cultural. As a result, designations of normal and abnormal genetic variation and normal and mutant genes do not furnish scientifically objective, acultural grounds for clinical intervention.

Chapter Five, "What's in a Cause?: The Pragmatic Dimensions of Genetic Explanations," departs from the concerns of the preceding three chapters over what counts as normal or abnormal genetic variation to consider the question of genetic causation generally. What is it to say that genes, whether normal or mutant, *cause* a trait? How are we to understand the phenomenon of geneticization — the increasing frequency with which differences among individuals are attributed to genetic differences? Chapter Five argues for a pragmatic account of genetic explanation. This is to say that when a disease or other trait is termed 'genetic,' the reasons for singling out genes as causes over other, also necessary, genetic and nongenetic conditions are not wholly theoretical but include pragmatic dimensions. Whether the explanandum is the presence of a trait in an individual or differences in a trait among individuals, genetic explanations are context-dependent in three ways: they are relative to a causal background of genetic and nongenetic factors; they are relative to a population; and they are relative to the present state of knowledge. Criteria like causal priority, nonstandardness, and causal efficacy that purport to distinguish objectively between genetic causes and nongenetic conditions either incorporate pragmatic elements or fail for other reasons. When the pragmatic dimensions of genetic explanations are recognized, we come to understand the current phenomenon of "geneticization" to be a reflection of increased technological capacities to manipulate genes in the laboratory, and potentially the clinic, rather than theoretical progress in understanding how diseases and other traits arise. This calls into question the value of

the search for theoretical definitions of designations like 'genetic disease' or 'genetic susceptibility' as directives for action.

This, then, is an overview of the content of the dissertation but, before moving on to Chapter Two, it is necessary to clear up some possible sources of terminological confusion. Thus far, I have referred to concepts like genetic variation, genetic normality, and genetic mutation in very general, potentially more or less evaluative, ways. These three concepts are closely related: genetic variation arises only as a result of genetic mutation (change) in the normal (preexisting) chromosomal structure; a gene is normal (functional) if it contains no mutations (defects); all normal genes have at some time in their evolutionary pasts been mutants; genetic variation and the mutations responsible for generating that variation make evolution through natural selection, and therefore the development of the adapted (normal) genome, possible; genetic variants may be considered to be normal, abnormal, or adaptively neutral. The dissertation as a whole looks at genetic variation and Chapter Two focuses on concepts of genetic normality and the senses of 'normal' that they express. The concept of mutation, however, is dealt with less extensively. Hence, it may be helpful to clarify 'mutation,' in somewhat more technical language before proceeding further. 'Mutation' has thus far been used in two ways: to refer to a change in chromosomal structure that may be beneficial, harmful, or adaptively neutral; and, to refer to a harmful allelic variant of a "gene."

'Mutation' was introduced to the study of heredity in 1901 by Hugo de Vries (*Die Mutationstheorie*) who was studying evolutionary mechanisms in *Oenothera lamarckiana* (evening primrose). De Vries used the term 'mutation' to describe sudden changes that arose in the plant's appearance, changes which he believed represented the formation of a new species and supported a theory of saltatory evolution.[5] However, with the sudden appearance of a white-eyed fly in one of T. H. Morgan's culture bottles in 1910, and the discovery that the mutant could be bred with normal flies, the concept of mutation came to include spontaneous heritable changes within a species. It was A. H. Sturtevant in

[5] It was later discovered that most of de Vries' "mutations" were actually genetic recombinants.

1913 who conceived that mutation could be responsible for producing a number of variant forms of a gene and gradually it was accepted that variant alleles are responsible for quantitative as well as qualitative variation in traits. In the early 1920s, H. J. Muller, who was successful in using radiation to induce mutations, came to understand these changes in terms of chemical alterations or physical damage to the chromosome (Bowler 1989). With today's knowledge of the mechanisms underlying DNA replication and conceptions of a genetic "code," 'mutations' as changes to chromosomal structure are understood largely as copy "errors" that occur accidentally in DNA replication (Ridley 1993, p. 32). For example, one base may be substituted for another or there may be an insertion or deletion of a single base. Changes that involve larger segments of chromosomes or even entire chromosomes may also be referred to as mutations: these include translocations (exchange of chromosomal material within a chromosome or between chromosomes), inversions, deletions, chromosomal fusions, and chromosomal duplications. Insertions of mobile genetic elements also yield changes in chromosomal structure. A definition of 'mutation' in 1976 as "any *hereditable* [*sic*] nucleotide base change, deletion, or rearrangement in the primary structure of DNA" (Siminovitch in Collins 1996, p. 256; italics mine) has been broadened today to include nonheritable changes, that is, changes in the DNA of somatic, as well as reproductive, cells.

Mutations can be beneficial, harmful, or neutral in their effects. Variation at the level of DNA may or may not be associated with variation at the protein level; variation in proteins, in turn, may or may not be associated with variation at the level of gross organismal phenotype. Point mutations may be synonymous (no change in the amino acid coded for) or nonsynonymous (change in the amino acid coded for); they may also be transitions (base substitution of a purine for a purine or a pyrimidine for a pyrimidine) or transversions (base substitution of a purine for a pyrimidine or vice versa). Insertions may or may not represent frameshift mutations that interfere with protein synthesis (Ridley 1993, pp. 73-74). Mutant alleles that are deleterious in their effects tend to be kept at low levels in populations due to negative selection. Mutations that are neutral or of fluctuating selective value may accumulate in populations. Consequently, an allele's frequency in a population may indicate its selective value. Alleles that occur at frequencies less than one percent are referred to as 'mutations' and are considered to be

deleterious; alleles that occur at frequencies greater than one percent are referred to as 'polymorphisms.'[6] Polymorphisms may be maintained in a population by different forms of balancing selection such as heterotic selection, frequency-dependent selection, or diversifying selection. Polymorphisms may instead be selectively neutral or slightly deleterious. Or, due to founder effect, a moderately or even severely deleterious variant may appear at a higher frequency in a small isolated population than it would in a large interbreeding population.

The distinction between mutations and polymorphisms is sometimes made instead on the basis of the effects of alleles in individuals rather than on their frequencies in populations. For example, the author of a textbook on "molecular medicine" writes: "If the presence of an abnormal allele causes the individual to have a disease, we say that the allele has a *mutation*. If an unusual allele does not cause any abnormality, we call the alternate form of the allele a *polymorphism*" (Ross 1992, p. 79). However, it is important to emphasize that the selective value of a particular variant depends on its genetic and environmental backgrounds. In Theodosius Dobzhansky's words: "A gene need not be unconditionally good or bad, useful or harmful, adaptive or unadaptive. If the environment changes, some genes that were favorable in the old environments may become unfavorable, and others may become favorable" (1962, p. 125). Nevertheless, some mutant genes are considered to be unconditionally bad. Such a belief inspired geneticist Herbert Spencer Jennings, in 1927, to write: "A defective gene — such a thing as produces diabetes, cretinism, feeblemindedness — is a frightful thing; it is the embodiment, the material realization of a demon of evil; a living self-perpetuating creature, invisible, impalpable, that blasts a human being in bud or leaf. Such a thing must be stopped wherever it is recognized" (in Paul 1995, p. 69).

---

[6] Specifically, Kimura (1982) writes: "polymorphism means coexistence of two or more allelic forms in a species, usually excluding the situation where the frequency of the most prevalent allele is higher than 99%" (p. 35).

Chapter Two

Defining Normality and Denying Variation: The Human Genome Project's "Presumably Representative" DNA Sequence

> [T]he sequence of the human DNA is the reality of our species. (Dulbecco 1986, p. 1056)
>
> The population geneticist, or the classical biologist, in defining the species, can point to a type specimen, an organism, and say that it exemplifies the species. The molecular biologist's view is that this organism is defined by its DNA. That DNA molecule can be sequenced to reveal the essential information that defines the type organism and hence the species. (Gilbert 1992, pp. 84-85)
>
> The Human Genome Project, which will create a stereotype of human genetic structure, is in a sense history's greatest exercise in platonic essentialism. (Weiss 1996, p. 1)
>
> The Definition of Man recited itself in my head: '... and each leg shall be jointed twice and have one foot, and each foot five toes, and each toe....' And so on, until finally: 'And any creature that shall seem to be human, but is not formed thus is not human. It is neither man nor woman. It is a blasphemy. (Wyndham 1958, p. 13)

Human diversity is immense. People come in all sorts of shapes and sizes, in a multitude of hues of hair, skin, and eye colour, in different gender configurations, of varied personalities, and with wide ranging abilities. While this is a surprise to no one, it is often forgotten that such diversity is not restricted to humans but is characteristic of much of the organic world. Amidst all the diversity, it is biology's task to search for unifying principles. Some biologists study variation itself, attempting to discern the relative importance of deterministic and chance factors in evolution or the processes by which speciation occurs. Evolutionary and population geneticists study variations in gene frequencies within and between populations, in space and over time. Other biologists ignore variation, seeking instead basic cellular and subcellular mechanisms universal to all living things. The laboratory approach of molecular genetics (and its classical and biochemical predecessors) seeks to discover universal principles that underlie hereditary transmission, gene action, and the structure and function of DNA.

When mice, flies, yeast spores, bacteria, or viruses are studied by laboratory geneticists, differences that appear irrelevant to the task at hand are ignored. The

particular is discounted in the search for the universal: "a" fly becomes "the" fly; the data collected on the laboratory population of flies is taken to represent flies everywhere; the "laws of heredity" discovered by studying flies are interpreted to hold for (at least diploid) organisms generally. A single fly potentially represents all flies in the strain, all flies in the species, all flies, all insects, all animals, or all organisms. The assertion of French molecular biologist Jacques Monod that "what was true for *E. coli* would be true for the elephant" (in Judson 1996, p. 592) exemplifies this approach. Evolutionary biologists and ecologists have long been critical of the "typological" and "essentialist" treatments of interspecific and intraspecific variation by their laboratory colleagues. Evolutionist critics of the HGP believe that its goal to produce, by the year 2005, "the complete sequence of a presumably representative human genome" (Maddox 1991, p. 11) epitomizes such "typological" and "essentialist" thinking. That "the" genomes of several "model organisms" (*E.coli*, *C. elegans*, yeast, mouse) will also be sequenced only compounds the problem.

Walter Gilbert, HGP proponent, Harvard molecular biologist, and co-developer of a major sequencing technology, makes no apologies for this approach: "Molecular biologists generally view the species as a single entity, sharply defined by a set of genes and a set of functions that makes up that entity" (1992, p. 84). He believes that we will come to understand ourselves as a species by identifying those DNA sequences in which we differ from nonmammals, nonprimate mammals, and nonhuman primates:

> At the end of the genome project, we will want to be able to identify all the genes that make up a human being. For example, we will compare the sequences of the human and the mouse and be able to determine the genes that define a mammal by this comparison, because the regions of DNA that code for protein are very well conserved over evolutionary time whereas the regions that do not have important functions are not well conserved. So by comparing a human to a primate, we will be able to identify the genes that encode the features of primates and distinguish them from other mammals. Then, by tweaking our computer programs, we will finally identify the regions of DNA that differ between the primate and the human — and understand those genes that make us uniquely human. (ibid., p. 94)

Closely associated with this typological and essentialist approach to species definition is the treatment of intraspecific variation as deviation from the species norm. The HGP assumes that "all the genes that make up a human being" can be defined in their "normal"

or "healthy" forms. The "representative" human genome represents not a typical or average human who differs in some finite number of DNA nucleotides from other primates or mammals but a normal or healthy human. For instance, it is claimed that the human genome sequencing project will allow "[t]he rapid and sure identification of genetic diseases" when DNA sequences of affected individuals are compared with the reference sequence (Maddox 1991, p. 12). James D. Watson, co-discoverer of the molecular structure of DNA, director of the Cold Spring Harbor Laboratory, and previous director of the National Institutes of Health's Genome Project, is eager for the sequencing project to be completed because he has spent his career "trying to get a chemical explanation for life, the explanation of why we are human beings and not monkeys" (1992, p. 164) and believes that "the genetic messages encoded within our DNA molecules will provide the ultimate answers to the chemical underpinnings of human existence" (1990, p. 44). Watson's yearnings are not existentialist. These "ultimate answers" encoded in the DNA concern questions of health and disease: "They will not only help us understand how we function as healthy human beings, but will explain, at the chemical level, the role of genetic factors in a multitude of diseases" (ibid.)

In this chapter, I evaluate the HGP's goal to produce, by 2005, "the complete sequence of a presumably representative human genome" in terms of how such a sequence may operate to define (perhaps, quite arbitrarily) a standard of genetic normality while denying the presence, prevalence, and propriety of genetic variation. The HGP aims to produce a single DNA sequence that is understood to represent not *a* mapped and sequenced human genome but *the* mapped and sequenced human genome; it is referred to as a representative, standard, or reference sequence. This is problematic in two, related, ways. First, it seems an implicit denial of the genetic variation that characterizes all biological species. In humans, we know that no individual is homozygous at all pairs of loci, that no two individuals except for monozygotic twins are identical at all loci, and that there is extensive genetic heterogeneity both within and between populations. As Theodosius Dobzhansky once noted: "the potentially possible genetic endowments are inexhaustible and a vast majority of them can never be realized.... It is, therefore, in the highest degree unlikely that any two persons (other than identical twins) ... have ever had, or will ever have, the same constellation of genes" (1962, pp. 30-31). Second, the

sequence may serve as a normative standard that treats variation as deviation, a situation with obvious medical and social implications. Although variation is typically treated as deviation from a norm by molecular and other laboratory geneticists, population geneticists, as the following passage from Walter Bodmer illustrates, are not immune:

> Analysis of normal human variability in facial features, character, and mental abilities is surely one of the real challenges of human genetics.... Knowledge of the total human genome sequence has profound implications ... for the better understanding of normal variation, and through that, hopefully making a contribution to solving the wider problems of society. (Bodmer 1986, pp. 12-13)

Bodmer is confident that normal and abnormal variation can be readily distinguished: abnormal variation is that which is associated with disease. However, *all* normal variation — in facial features, character, and mental abilities — does not appear to be equally acceptable. Bodmer's remark that a better understanding of normal variation could contribute to solving "the wider problems of society" leads one to conclude that "normal variation" can itself be parcelled into variation that is, if not "normal" or "abnormal," at least "good" or "bad."

I begin the chapter by investigating the HGP's treatment of genetic variation. I point out that genetic linkage mapping and physical mapping do not themselves deny the presence, prevalence, and propriety of genetic variation. Genetic maps, with their thousands of polymorphic markers, are, in fact, testaments to the vast extent of genetic variability within the species. As is consistent with the history of laboratory genetics, human molecular genetics uses variation as a tool to elucidate "normal" structure and function. Guided by the hypervariable markers displayed on genetic maps, gene mappers locate disease genes; armed with probes and supplied with DNA from the genomic libraries organized by physical maps, gene hunters clone normal and disease genes alike. Chapter Four addresses questions surrounding gene maps and the distinctions between normal and abnormal genetic variation and normal and mutant genes. In this present chapter, I am concerned with the maps specific to the original aims of the HGP: the production of genetic linkage maps, physical maps, and the "ultimate" sequence map. The presence, prevalence, and propriety of human genetic variation appears to be ignored or denied, if not by the genetic and physical maps, then certainly by this "ultimate" goal — a single human DNA reference sequence against which other sequences will be compared.

In an attempt to grasp how a single DNA sequence might represent humanity in all of its diversity, I look at how biologists have used the human mitochondrial DNA (mtDNA) reference sequence over the past fifteen years. This reveals an evident tendency for researchers to treat the mtDNA reference sequence as a standard of genetic normality in both statistical and functional senses of 'normal.' "Wild-type" and "consensus sequence," the prevalent concepts of normality in twentieth-century genetics, similarly convey both of these senses. Although different notions of "normal" — what is typical, functional, or ideal — are often conflated in everyday usage, these have been distinct concepts in biology since the nineteenth century. I discuss this with reference to the writings of Claude Bernard, Adolphe Quetelet, and Francis Galton, as well as to the quite extensive recent literature on the rise of probabilistic thinking in nineteenth-century science.

2.1 Genetic Maps, Physical Maps, and the "Ultimate" Sequence Map

As outlined in Chapter One, the scientific goals of the HGP are to produce: first, a genetic map of the human genome; second, a physical map of the human genome; and third, the complete DNA sequence of the human genome. I argue in this section that, of these three, only the sequence map has the potential to serve as a standard of genetic normality that denies the presence, prevalence, or propriety of human genetic variation. Genetic and physical maps are properly regarded as tools for use in gene mapping and genome sequencing. They are not themselves representations of the normal and the pathological. Genetic maps, in fact, are testaments to "normal" human variability. I begin this section by explaining genetic and physical maps: how they are compiled; what they represent; and how they are used as tools in gene mapping and genome sequencing. I attempt to understand how the HGP's "representative" genome sequence may, taken as a reference sequence, serve to institute a standard of genetic normality.

2.1.1 Genetic Maps

The genetic maps associated with the HGP continue the tradition of chromosome linkage mapping established early this century by T. H. Morgan and his outstanding student A.

H. Sturtevant. In 1910, Morgan discovered that white eyes in *Drosophila melanogaster* is a "sex-limited" (what is now called a "sex-linked") trait. In 1913, Sturtevant constructed the first genetic linkage map, ordering six traits along the X-chromosome by marking them with horizontal lines across a vertical line (representing the chromosome) according to the relative distances separating them. Linkage mapping makes use of the frequency of crossing-over of alleles that occurs between maternally and paternally inherited chromosomes during meiosis. The likelihood that two genes will be inherited together is proportional to their proximity to one another on the chromosome since this proximity makes recombination less likely to occur. Conversely, the probability of their separation is proportional to their distance apart. Therefore, the frequencies with which different genes do recombine provide an estimate of the distance between them. This distance is not actual but relative; whereas distances on physical maps are expressed in number of nucleotide bases, distances on genetic maps are expressed in centiMorgans (cM). One cM is equal to a one percent chance that two genetic markers will be separated during meiosis by recombination.

Genetic linkage mapping can be carried out only if variation is present at the relevant gene loci. Classical geneticists could discern genetic variability only at the level of gross organismal phenotype — eye colour or wing shape in *Drosophila*, for example. The first human trait was mapped in 1911 when E. B. Wilson located colour-blindness on the X-chromosome by discovering that colour-blindness and male sex are phenotypes that segregate together (McKusick 1986). It was not until 1968 that a human trait was assigned to a specific autosome. Researcher R. P. Donahue discovered a heteromorphism of chromosome-1 in his own family and found it to segregate with the Duffy blood group (ibid.). Since the days of classical genetics, technological developments have permitted genetic variability to be discerned at more basic levels than gross organismal phenotype. Molecular phenotypic traits such as blood protein types became accessible to gene mappers using the techniques of chromatography and electrophoresis from the mid-1950s on, allowing linkage between gross phenotypic traits and molecular traits to be determined. In 1970, chromosomal staining techniques that reveal characteristic banding patterns for each human chromosome became available (Judson 1992, p. 69). Gross structural chromosomal variations due to large inversions, translocations, deletions, and

duplications could be identified and sometimes linked to phenotypic traits of interest. Chromosomal staining, combined with the technique of somatic-cell hybridization published by Mary Weiss and Howard Green in 1967, meant that proteins detected in cell cultures containing hybrid mouse-human cells could be mapped to specific chromosomes or parts of chromosomes (ibid., pp. 68-69). This permitted a molecular trait to be mapped directly to a chromosome without needing to rely on its linkage to another phenotypic trait, whether molecular or gross organismal.

The advent of molecular techniques capable of detecting structural variation at the level of the genome has made it possible for linkage mapping to proceed even where genetic variation is not associated with gross changes in chromosomes or variation in molecular or gross organismal traits. A phenotypic trait of interest that segregates with a defined genetic marker can be mapped to a specific region of the genome. Genetic markers are hypervariable loci, chosen for the likelihood that they will differ, or be heterozygous, in any two copies of a chromosome, whether in the same or in different individuals. The polymorphisms used in genetic mapping tend to be situated in regions of the genome subject to minimal functional constraint. Extensive amounts of genetic material believed to be nonfunctional or redundant are present in eukaryotic genomes. This "junk" DNA, which is believed to comprise 90-95 percent of the human genome, has proved to be of great assistance in the construction of dense genetic maps. The more densely that genetic markers cover a genetic linkage map, the more likely it is that a marker will be found which demonstrates close linkage to a disease gene that is being sought. The 1996 Généthon map is anticipated to offer linkage mapping for "monogenic" diseases to be carried out to the centiMorgan level in most cases (Dib et al. 1996). It is also hoped that dense genetic maps will contribute to the identification of genes involved in "complex" or non-Mendelian patterns of inheritance — quantitative traits as well as qualitative traits that involve genetic heterogeneity, incomplete penetrance, and gene-gene and gene-environment interactions (Lander and Botstein 1986).

Genetic linkage maps, influenced by technological developments along the way, have been constructed using a variety of types of genetic markers. The first global human genetic map, published in 1987 by Helen Donis-Keller's group at Collaborative Research, used restriction fragment length polymorphisms (RFLPs) as markers. RFLPs arise due

to single base differences that either create or eliminate recognition sites when sample DNA is exposed to a variety of restriction enzymes. RFLPs are detected by a technique called Southern blot analysis. The 1987 Collaborative Research map comprised some 400 RFLPs covering an estimated 95 percent of the genome (*Nature Genetics* 1994 editorial). RFLPs subsequently came to be replaced as genetic markers by polymorphisms that are detectable using the polymerase chain reaction (PCR) and can serve as landmarks for both genetic and physical maps. Jean Weissenbach's Généthon group published the first genetic map using only PCR-detectable markers in 1992; this was a precursor to their comprehensive genetic map of the human genome that was published in the 14 March 1996 issue of *Nature* (Dib et al. 1996). The Généthon genetic maps use a type of microsatellite marker (also called variable number of tandem repeat polymorphisms [VNTRs], simple sequence length polymorphisms [SSLPs], or short tandem repeat polymorphisms [STRs]): specifically, polymorphisms that are highly variable with respect to the number of dinucleotide repeats — $(AC/TG)_n$ — that are present. Microsatellites are preferable to RFLPs as markers for genetic maps, not only because they can be detected by PCR, which is easier and quicker than the detection of RFLPs using Southern blot analysis, but because they are more polymorphic than RFLPs, ubiquitous in eukaryotic genomes, and found abundantly throughout the genome (every 50,000 base pairs, on average) (Neilan et al. 1994). Very recently, researchers have begun to compile single-nucleotide polymorphisms (SNPs), which are common alterations in a single nucleotide along a stretch of DNA, in order to construct a genome-wide SNP map that will be useful in identifying genes that contribute to complex traits (Collins et al. 1997).

As representations, the genetic maps associated with the HGP do not deny the presence, prevalence, or propriety of human genetic variation. These genetic maps are tools to be used for gene mapping. Since Sturtevant's first map in 1913, genetic maps have served in this role. As increasing numbers of trait loci are mapped, the markers on the map become increasingly dense, and there is an increasing likelihood that additional trait loci will be successfully identified. It is something new, though, for genetic linkage maps to be developed for the sole purpose of being used as tools for gene mapping, as is the case for the HGP. Past genetic linkage maps have also served as representations of the normal and the pathological. To be more accurate, they have overwhelmingly been

representations of the pathological. As mentioned already, linkage mapping cannot be carried out unless variation is present at the relevant genetic loci and the technical means exist to detect that variation. Until recently, gene loci could be mapped only if variant alleles at these loci were associated with phenotypic variation. For the most part, it has been "mutant" or disease phenotypes that are mapped. Linkage maps therefore have been "mutant" or "morbid" maps: "from the start the genetic map of any species — molds or flies, maize or humans — has been primarily the map of defects" (Judson 1992, p. 47). By contrast, as representations, the genetic linkage maps associated with the HGP are testaments to the "normality" — the presence, prevalence, and propriety — of genetic variation in the human species. Each of the 5264 markers located to 2335 positions on the 1996 Généthon genetic map (Jordan and Collins 1996, p. 111) represents a site in the human genome that is highly variable within and between individuals.

Although the genetic maps associated with the HGP have been characterized as "a basic description of the structure of the human genome" (White et al. 1986, p. 29), they are generally regarded as tools to be used in gene mapping. The polymorphic loci represented on genetic maps are unlikely themselves to be functional, although it is possible that some small number of them may be found to influence variation in quantitative or complex phenotypic traits. But as tools for gene mapping, as "basic structural descriptions of the genome," as testaments to human genetic variation, the question arises whether genetic maps are variable enough. The genetic material used to create the maps comes from a limited number of individuals of particular ethnic backgrounds. The 1996 Généthon genetic map (Dib et al. 1996) uses material from the collection maintained at the Centre d'Etudes du Polymorphisme Humain (CEPH) in Paris. Specifically, DNA from eight CEPH families (134 individuals) was used for all of the chromosomes except the X-chromosome for which the DNA of 20 families (304 individuals) was used. CEPH was set up in 1984 to act as a depository of cellular material. The material could be used by any international group of researchers provided that groups share findings that are then stored in a public database maintained by CEPH. As of 1994, cell lines from sixty large families from France, Pennsylvania, Utah, and Venezuala were stored at CEPH (Wilkie 1994, p. 90). The CEPH families are well-characterized with extensive pedigree information available to researchers: included are

the Utah Mormon families studied by Raymond White's group at the University of Utah, Amish families from Pennsylvania in whom genetic studies of manic-depression have been carried out, the Venezuelan families used to map the Huntington's gene to chromosome-4, and the French families who have contributed to efforts to map the muscular dystrophy gene. It is not so much a concern for genetic maps that these families were of interest to researchers because of the incidence of specific hereditary diseases in their families; rather, since hypervariable genetic markers, not genes, are being mapped, the important issue is whether DNA has been sampled from an adequate number of families of diverse ethnic and racial backgrounds for the genetic linkage maps to be useful tools for gene mapping worldwide.

2.1.2 Physical Maps

Physical maps order collections of actual genomic fragments by chromosome. These collections are called "libraries." Genomic libraries are "built" by using restriction enzymes to fragment many copies of a single genome into chromosomal fragments. The complete physical map of the human genome will consist of a collection of ordered overlapping large fragments of recombinant (cloned) genomic DNA for each chromosome. Any individual nucleotide base would be contained in at least one clone (Little 1990, p. 611). These contigs — contiguous overlapping fragments — are attached to viral "vectors" or otherwise inserted into the chromosomal machinery of a variety of microorganismal hosts. Phage libraries contain bacteriophage (virus) clones with attached human DNA; cosmid libraries contain inserts of human DNA in bacterial plasmids; yeast artificial chromosome (YAC) or bacterial artificial chromosome (BAC) libraries contain yeast or bacterial cells with fragments of human DNA attached to the remaining nub of a yeast or bacterial chromosome. Genetic maps assist in the ordering of these fragments by providing markers as the framework or "ordered scaffold" upon which physical maps can be constructed. Techniques such as restriction mapping and *in situ* hybridization are used to order clones by determining which fragments have overlapping segments.

Physical markers are chosen, not for the likelihood that they will vary within and between individuals as for genetic markers, but for their uniqueness, the likelihood that

they occur only once in the genome of any individual. Useful physical markers include expressed sequence tags (ESTs) from cDNAs and sequence-tagged sites (STSs). STSs are generated by PCR primers and amplify just a single chromosomal site (Rowen et al. 1997, p. 605). It is possible to convert microsatellite markers on genetic maps into STSs which permits the integration of physical and genetic maps. Markers are localized using genetic mapping, fluorescence *in situ* hybridization, and radiation hybrid mapping (ibid.). Physical maps are useful for gene identification in several ways: once genetic mapping localizes a gene to a region of a chromosome, physical maps locate possible candidate genes and appropriate clones for sequencing; cDNA obtained by reverse transcription from mRNAs can be physically mapped by *in situ* hybridization of a cDNA probe to a genomic library; genes can be discovered by sequencing cloned DNA fragments ordered by physical maps. With STSs and PCR technology available, researchers are able to amplify DNA from laboratory samples and no longer need to maintain DNA clone libraries. Hence, STS markers, if they are to be useful across numerous individual genomes, should not be highly variable in a population. However, high resolution physical maps of the genome *and* genomic clone libraries are absolutely necessary for large-scale DNA sequencing and the eventual compilation of the "ultimate" sequence map.

One recent physical mapping effort (Chumakov et al. 1995) uses yeast artificial chromosomes (YACs). The map is estimated to cover about 75 percent of the genome using 225 contigs that have an average size of about 10 Megabases. The framework for the physical map is provided by STSs from the Généthon 1993-94 linkage map. YAC clones of an average length of one Mb from a large collection ("library") were screened to establish "links" or potential overlaps between them. Several screening procedures were used: STS screening (using STSs from Généthon's genetic map), screening by hybridization, and fingerprinting. Using these three types of "links," contigs could be assembled spanning the intervals between adjacent genetic STS markers. The 1995 CEPH-Généthon physical map was made available on the Internet and primary copies of the CEPH YAC library were sent to eight genome research centres worldwide in order to permit the distribution of clones. It appeared that the CEPH-Généthon physical map and the CEPH YAC library would furnish the map landmarks and the source materials to generate a significant portion of the sequence map. The CEPH YAC library is

composed of 98,208 YAC clones that represent about 17 genome equivalents in total. These were derived from a single human male lymphoblastoid cell line, Boleth (Chumakov 1995, p. 176). Thus, the "ultimate" DNA sequence map produced early next century would be a composite of sequenced DNA taken from a fairly small number of individuals, the Boleth donor among them. However, recently, a mandate was issued by the NIH and DOE instructing centres involved in large-scale sequencing to use only clone libraries obtained from anonymous donors who gave proper consent. This is in order to prevent the possibility of future genetic discrimination against the donor and her or his relatives. It means that sequencing centres must replace almost all of the existing clone libraries and physical maps (Rowen et al. 1997, p. 606).

2.1.3 The "Ultimate" Sequence Map

The genetic and physical maps are the first two steps in the three-part HGP plan that culminates in the "ultimate" map: that of the complete sequence of "the" haploid human genome. Physical maps are a key step in producing the complete sequence; they break the genome into "manageable chunks" which can then be sequenced (Little 1990, p. 611). Sequence-tagged sites (STSs) are markers on the physical map that serve as "replicable milestones in otherwise unknown territory" (Maddox 1991, p. 14). Typically, 40- to 200-kilobase segments of DNA, each represented by a single bacterial or yeast clone, are prepared and subcloned. A large centre will have 500 such clones in intermediate stages of the sequencing process at any given time. Completed consensus sequences are annotated and submitted to the public database. Annotated information may indicate possible errors or include functional information — the presence of a gene, regulatory region, etc. There exists an internationally agreed upon aim that the human genome sequence will be finished to a high degree of accuracy (99.99%) (Bentley 1996).

To aim for a DNA reference sequence that is 99.99 percent accurate refers only to accuracy in the sequencing process. It means that there is no contamination by non-human DNA, for example, yeast or bacterial DNA from the cloning process. It means that the sequence is ordered correctly and that there are virtually no errors with respect to the actual DNA sample that was sequenced. There are few strictures, however, placed on what constitutes a suitable or acceptable reference sequence. A reference sequence

may be a sequence derived from a single individual or it may be a composite derived from the DNA of multiple individuals. A reference sequence may or may not be a consensus sequence; a consensus sequence is obtained by comparing multiple DNA sequences and assigning to each nucleotide position the base that occurs most frequently among these sequences. There are two different senses of 'consensus sequence' implicated here. In the last paragraph, I wrote: "Completed consensus sequences are annotated and submitted to the public database." 'Consensus sequences' in this sentence refers to the generation of a single sequence from the comparison of multiple sequences all of which derive from the *same* individual. The aim is to ensure accuracy in the actual DNA sequencing procedure. On the other hand, when a consensus sequence is obtained by comparing the DNA sequences of a multiple number of *different* individuals, the aim is to produce a reference sequence that is accurate with respect to the population it is supposed to represent.

The HGP's "presumably representative sequence" will not be a consensus sequence of this latter type. It will be a composite sequence that is compiled from sequence data obtained from a relatively small number of individuals. The sequence in a particular region of the map is likely to have been obtained from a single individual. The significance of this depends on what we understand the reference sequence's representational status to be. If the reference sequence is to represent merely some typical human individual, presumably, DNA sampled from any random individual would do. Humans are not easily confused with members of species that are our closest primate relatives. We are all more or less healthy, more or less intelligent, more or less attractive, more or less able-bodied, etc. A composite genome reference sequence that is constructed by connecting together partial genome sequences from a number of individuals is warranted on this view because it represents some average individual who *could* feasibly exist without belonging to any actual individual. This appears to be Victor McKusick's position:

> The question often asked, especially by journalists, is "Whose genome will be sequenced?" The answer is that it need not, and surely will not, be the genome of any one person. Keeping track of the origin of the DNA that is studied will be important, but the DNA can come from different persons chosen for study of particular parts of the genome. Such an approach is consistent with that of most

> biologic research, which depends on a few, and even on single individuals, to represent the whole, and with the fact, well recognized by geneticists, that there is no single normal, ideal, or perfect genome. (1989, p. 913)

Although geneticists may recognize that "there is no single normal, ideal, or perfect genome," the HGP's "presumably representative sequence" is also a "presumably" healthy one. If the reference sequence is to serve as a genetic standard of normal functioning and health, the source of the material that is sequenced and how the sequence is cobbled together becomes more important than were it merely to represent some typical or average genome. There is no guarantee, for instance, that "representative" individuals presumed to be healthy may not at a later date be found to suffer from some hereditary illness. The particular alleles represented by the sequence will be a function of their frequency in the population and will therefore depend on the ethnic and racial backgrounds of the individuals whose genomes are sequenced.

This notion of a reference sequence as a genetic standard of normal functioning and health is consistent with how reference sequences for individual genes are currently understood in biomedical research. Take, for example, one research group's attempts to establish the normal nucleotide sequences of the α- and β-chains of pyruvate dehydrogenase (E_1), which is an enzyme that acts in the liver to break down pyruvate (Ho et al. 1989; Ho and Patel 1990). Individuals with depressed levels of E_1 activity suffer from lactic acidosis and varying degrees of neurological impairment. These researchers were frustrated in their attempts to identify mutations in either $E_1\alpha$ or -β mRNAs taken from such individuals because reliable reference sequences were unavailable. Hence, they undertook the task of establishing "authentic" and "unambiguous" cDNA/mRNA E_1 reference sequences. The $E_1\alpha$ cDNA reference sequence was obtained by sequencing a 1423 base pair cDNA clone from a human liver cDNA library (Ho et al. 1989). The "authenticity" of the sequence was "validated" in several ways. The sequence was found to be identical in its 1362 bp overlap with a smaller $E_1\alpha$ cDNA clone from the same cDNA library that had been sequenced previously (Wexler et al. 1988). Since these DNA clones originated in a single individual, this represents a test of the accuracy of the actual sequencing procedure. Additionally, the sequence was identical to those of three overlapping cDNA clones covering its entire length that were generated using reverse

transcriptase and PCR amplification applied to liver and skin fibroblast mRNAs from several different sources. This, then, is a test of the validity of the reference sequence as a representation of the population generally. The $E_1\alpha$ cDNA reference sequence is a consensus sequence — all of its elements are shared by more than one cDNA clone from different individuals. The $E_1\beta$ cDNA reference sequence, on the other hand, is a composite sequence derived from two cDNA clones from the same human liver cDNA library that did not overlap in their entirety (Ho and Patel 1990). For $E_1\alpha$ and -β reference sequences alike, the authors were confident that they had taken the necessary measures to ensure accurate results. They explained differences between the $E_1\alpha$ reference sequence and three previously published sequences, and between the $E_1\beta$ reference sequence and one previously published sequence, as due to: tissue specificity since mRNAs were obtained from skin fibroblast as well as liver cells; "peculiarities" of other samples; and cloning or sequencing artifacts in other labs.

But is there a single sequence that characterizes a "normal" gene or a "normal" genome and justifies these researchers' search for a completely "unambiguous" and "authoritative" sequence and their presumption that each and every nucleotide of the sequence can be known with accuracy and is otherwise an "error?" 'Error' here has two senses: it represents either our failure to discern the correct nucleotide or a mutation to that nucleotide. What is being sought, after all, is "the unambiguous reference sequence needed for the characterization of genetic mutations in pyruvate dehydrogenase-deficient patients" (Ho et al. 1989, p 5330), that is, sequences that "accurately represent normal human $E_1\alpha$ and -β mRNAs" (Ho and Patel 1990, p. 298). The availability of normal DNA and RNA sequences will help not only to detect mutations in the clinic but to provide the means for learning about normal structure-function relationships in the laboratory. Likely, there is a great degree of functional constraint exercised on transcribed regions of the genome, as represented by these cDNA clones. Yet, studies of amino acid polymorphisms in different proteins have revealed that many variant sequences are fully functional. Add to this silent nucleotide substitutions that do not result in any amino acid changes and it becomes even more evident that there is not a single normal sequence for a given transcript. Apart from transcribed regions of the genome there is additional variability in regulatory regions and, especially, in areas of indiscernible

function. Insofar as it is questionable whether an individual gene can be characterized by a unique reference sequence, how are we to understand the HGP's aim to produce a composite reference sequence for the entire genome? What kind of standard will such a reference sequence represent? A look at how "the" human mitochondrial genome reference sequence has been used since it was produced in 1981 might be useful.

2.2 The Human Mitochondrial DNA Reference Sequence: A Case Study

"The" mitochondrial genome was completely sequenced in 1981 by Frederick Sanger's group at the Medical Research Council (MRC) Laboratory of Molecular Biology in Cambridge, U.K. (Anderson et al.). Sequencing was carried out on mtDNA obtained from two sources: the placental tissue of a single (white European) individual and the HeLa cell line. Apparently, in several places where sequencing difficulties were encountered, sequence data from bovine mtDNA was substituted. The single mtDNA sequence produced by these researchers is now commonly referred to as the "Cambridge sequence" or the "Cambridge reference sequence (CRS)." Compared to the estimated three billion nucleotide base pairs present in the human nuclear haploid genome, the mitochondrial genome (which is haploid because it is inherited only from the mother) is a paltry 16,569 base pairs. The mitochondrial genome reference sequence, when first printed, filled three closely printed pages of *Nature*. Printing the complete nucleotide sequence of a single haploid nuclear genome, by contrast, would require the equivalent of about 13 sets of the *Encyclopedia Britannica* (McKusick 1986, p. 24).[1]

---

[1] Besides size, there are significant differences between nuclear and mitochondrial genomes. Most of these I do not think interfere with an analysis of the bearing of a published reference sequence on our understandings of genetic variation and genetic normality. Mitochondrial DNA (mtDNA) has a higher mutation rate and is more rapidly evolving than nuclear DNA; mtDNA is present in the cytoplasm and therefore maternally inherited; there is no meiotic recombination between maternal and paternal chromosomes as for nuclear DNA; since cells contain thousands of mtDNA molecules there can be a mixture of normal and mutant forms present (heteroplasmy) to varying degrees within different tissues of the same individual, as opposed to the limited qualitative options of homozygosity or heterozygosity in nuclear DNA; the genetic code in mtDNA differs from the so-called "universal" code. mtDNA's high mutation rate and the consequent better odds on encountering sequence polymorphisms, may even facilitate finding out

Mitochondrial genome sequence data is used by researchers in many different ways and to investigate all sorts of different biological, medical, and anthropological questions. Sometimes the sequence merely furnishes the coordinates that permit nucleotides to be numbered so that the sequences of different genomes can be compared and the results communicated to others. Knowledge of the characteristic sequence of nucleotides in a region also allows primers to be made so that specific segments of different genomes can be cloned, sequenced, and compared. In addition, the known locations of restriction sites for the reference sequence facilitates the construction of unambiguous restriction maps and the comparison of restriction fragment length polymorphisms (RFLPs) in different individuals. In these cases, the sequence and RFLP data of sampled individuals are compared amongst each other, and not to the reference sequence. For example, in some disease studies, the mtDNA of affected family members is compared to that of unaffected members, or mtDNA from tumour tissue is compared to that of healthy tissue (Heerdt et al. 1994), or the mtDNA of those with a disease suspected to be of mitochondrial origin is compared with that of a healthy control group. In population studies, mtDNA variation may be compared within and between populations to make inferences about the evolutionary past — for example, the "Out of Africa" hypothesis, that postulates a single African woman who lived 200,000 years ago as the last common ancestor of modern human mtDNA (Cann et al. 1987).

Frequently, the reference sequence is used not just to provide coordinates, primers, and RFLP sites, but for direct comparison with sample sequences. This occurs in two different ways. One way is by treating the Cambridge sequence as just one of the many

---

how a single sequence of nucleotides can be representative of an entire species and yet accommodate the genetic diversity that is characteristic of all species. I do have reservations, however, concerning three aspects. The first is that the mitochondrial genome is compactly organized; compared with the nuclear genome, a vastly greater percentage of it is functional. Except for the D-loop region, there are no or very few noncoding bases found between adjacent genes (Anderson et al. 1981, p. 457). Whereas over 90% of the mitochondrial genome codes for proteins, at least this much of the nuclear genome is noncoding. The second is that there are relatively few genes compared with the nuclear genome, and therefore phenotypic complexity is much less. The third is that since mtDNA is maternally inherited and there is no recombination, siblings (even half-siblings with the same mother) will differ in mtDNA sequence only as a result of mutations that arise in the oocyte or during development, or due to varying amounts of heteroplasmy.

available mtDNA sequences stored in databases. Forensic investigations may estimate the significance of a match between samples of mtDNA taken from human remains and from those suspected to be maternal relatives by comparing the haplotype to a large number of sequences, the Cambridge sequence among them (Stoneking & Melton 1995, p. 10). Evolutionary studies have used the Cambridge sequence, along with the published sequences of other humans as well as nonhuman primates, to estimate the date of the deepest root of the mtDNA tree for humans (Hasegawa & Horai 1991). This phylogenetic approach may take the Cambridge sequence to be a representative European sequence that can meaningfully be compared to representative African and Asian sequences and those of nonhuman primates to establish the African origins of modern humans (Horai et al. 1995), or even to be a representative human sequence that can be compared with the sequences of other primates and nonprimates to resolve the African ape trichotomy into two evolutionary lineages leading separately to gorillas and to humans and chimpanzees (Ruvolo et al. 1991; Horai et al. 1992).

The other way the Cambridge sequence is used for direct comparison with sample sequences is not just as one sequence among many, but as the baseline reference against which all others are compared. Forensic identification may proceed by using the Cambridge sequence as the basis for identifying polymorphisms in mtDNA samples obtained from the available remains and from suspected maternal relatives, and then establishing the frequency of such a match in the population (Ginther et al. 1992, p. 137; Holland et al. 1993, p. 549). Divergences from the Cambridge reference sequence also define the maternal lineages and haplotypes that are used in population studies to make inferences about the evolutionary relationships among human ethnic groups (Horai and Hayasaka 1990), the patterns of migration of early humans (Sykes et al. 1995), and the genetic diversity and likely origins of local populations (Ward et al. 1991; Côrte-Real et al. 1996[2]), as well as to identify the ethnic backgrounds of individuals (Torroni & Wallace 1994). Clinical studies of diseases of possible mitrochondrial origin usually

[2] The Côrte-Real et al. study calculates the percentage of "Cambridge Reference Sequence (CRS)" haplotypes (for a 302 bp region of the control region) found in each of several populations of the Iberian peninsula and finds it to be the most prevalent, ranging from 13.3% in Andalusia and Northern Spain to 26.2% in the Basque (1996, p. 334).

begin by comparing mtDNA taken from affected individuals to the reference sequence. Any variants identified in affected individuals are then investigated to find out if they are linked to the disease. Variants are labelled mutations, and added to the clinical mutations data set, if they are present in subjects with the disease but not in control subjects; if nucleotide differences are associated with amino acid changes in highly conserved coding sequences; or if maternally related family members, whether affected or unaffected, are found to be heteroplasmic, that is, a mixture of both mutant and wild-type mtDNA is present (Nishino et al. 1996). If none of these eventualities holds, the identified variants are suspected to be neutral and are added to the list of known polymorphisms (Heerdt et al. 1994). However, where a polymorphism is present in both disease and control subjects, but the incidence is higher in the former than in the latter, it may be considered to be a mildly deleterious mutation that, in combination with other factors, increases susceptibility to disease (Nakagawa et al. 1995). Haplotype analysis may also contribute to disease studies. It is easier to determine whether variants identified by comparing the mtDNA sequences of affected individuals with the reference sequence are disease mutations or population-specific polymorphisms if the unaffected individuals in the control group are of similar ethnic background (Jun et al. 1994; Torroni & Wallace 1994).

In the reference sequence's various uses, in what ways is variation denied and normality defined? In no way, I think, is the presence and prevalence — the fact — of genetic variation denied. Forensic identification is possible only because individuals are genetically similar to their relatives and dissimilar to others. Likewise, both similarities and differences in mtDNA haplotypes contribute to evolutionary, anthropological, and disease studies. The Cambridge sequence has been used with other sequences to estimate the extent of nucleotide diversity (Aquadro and Greenberg 1983). MITOMAP,[3] the human mitochondrial database available to the general public on the World Wide Web, maintains a database compiled from all published research on mtDNA that includes two types of variants alongside the "standard" sequence: the "clinical mutation data set" and the "population variation data set" (Kogelnik et al. 1996, p. 177). The clinical mutation

[3] It is hoped that MITOMAP will serve as a model for the development of information storage and retrieval systems for the rest of the human genome.

data set provides information on the nucleotide base substitutions (point mutations) and the rearrangements (deletions and insertions) associated with disease. The population variation data set identifies known polymorphisms and their distributions in different populations. Hence, the availability of a reference sequence need not deny variation; the sequence may even function as a tool by means of which the full extent of variation can be elucidated. The problem instead concerns how genetic variation is to be regarded — its value. Is variation to be understood as difference or as deviation? This depends, at least in part, on the authority that attaches to the reference sequence in its various uses.

In each of these uses, the authority of the reference sequence *can* be minimal; the sequence regarded as being standard, normal, or representative in the purely descriptive sense. Where the Cambridge sequence is used to provide coordinates, primers, and RFLP locations, it is a reference only in the sense that it provides a convenient shared framework, that in itself has little meaning. It is merely a structure or grid adopted by convention to permit numbers to be attached to nucleotide positions so that researchers can communicate their findings to one another and a data base of these findings maintained. Where the Cambridge sequence is one of many sequences available for pair-wise comparison, it is accorded no special status since all of these are being used as reference sequences. Even when it is taken to be "representative" of the entire human species (or of European humans), this can be understood in the most basic arbitrary way in which a sequence obtained from *any* human would be representative merely in virtue of this individual being human (or European). This perspective characterizes phylogeny and population studies where there is worry about chimeric sequences contaminated by another species or subspecies, and sequences established from one individual are therefore preferred.[4] The Cambridge sequence is criticized because, although most of it was obtained from a single (European) individual's placental tissue, mtDNA from the HeLa cell line derived from an African-American was also used, and some ambiguous nucleotides were designated to be the same as for bovine mtDNA that had been sequenced by the same group (Ozawa 1995, p. 182).

---

[4] A group who sequenced the mtDNA of a single gorilla writes: "The sequence was established from one individual and thus nonchimeric" (Xu & Arnason 1996, p. 691).

Where the Cambridge sequence is used as a baseline reference against which other sequences are compared, differences need be only that, not deviations or abnormalities. Here, the reference sequence is taken to be "representative" in a somewhat more stringent, but still descriptive, sense, in that it is typical of the population or species. For this reason, some researchers have emphasized the importance of establishing whether nucleotides in the Cambridge sequence that have rarely or never been otherwise observed are errors or infrequent variants (Howell et al. 1992) and others have used an "edited" version of the Cambridge sequence in their work (Horai et al. 1995). While haplotypes defined by the Cambridge sequence do seem to occur frequently in European populations, there was no guarantee that this would be the case and there is no basis for assuming the Cambridge sequence to be typical of all, especially non-European, populations. For this reason, Marzuki et al. (1991, 1992) propose that a "consensus sequence" replace the Cambridge sequence as the "best reference base for the study of human mtDNA variants" (1991, p. 142). Although the Cambridge sequence has been useful, they fault it for being a "composite" of human placental and HeLa cell mtDNA rather than being a "consensus" of mtDNA sequences taken from different individuals. Consensus sequences aim to be standard, normal, or representative in the descriptive sense of these terms. Consensus sequences are established by comparing sequenced DNA taken from two or more individuals and including, at each position in the sequence, the nucleotide that occurs most frequently at that site. A consensus sequence may be for the entire genome, an individual gene, or even a nonfunctional repetitive DNA sequence. Statistical significance, of course, depends on the number of genomes sampled and the range of populations from which the samples originate (the Marzuki consensus sequence was derived from thirteen unrelated mtDNA sequences).

However, either of these reference sequences, whether the Cambridge sequence or a consensus sequence that replaces it, seems destined to amass authority in the slide from descriptive to evaluative senses of 'standard,' 'normal,' or 'representative.' Although the Cambridge sequence numbering furnishes a grid researchers share, it is also the authoritative reference on mitochondrial genome size. Very seldom does an author write: "Human mitochondrial DNA [is] a circular molecule of about 16,500 base pairs" (Wainscoat 1987, p. 13). Rather, one usually reads: "*the* mitochondrial genome is 16,569

base pairs," with the 1981 article provided as reference. The authority extends further. The Cambridge sequence often determines what counts as a mutation: "When the different sequence from the published one was observed, the frequencies of the mutation both in the whole patients and controls were screened" (Nakagawa et al. 1995, p. 665).[5] It may be suggested that the Cambridge sequence is a standard only by convention and has no normative content, that is, it in no way indicates what is frequent or functional. But insofar as the Cambridge sequence determines what counts as an insertion or deletion in other sequences, it defines the normal length of mitochondrial genomes. Mutations, whether deletions, insertions, or substitutions, are never properties of the reference sequence, but only of those sequences compared to it. To insert, to delete, to substitute: all of these terms express the modification of some preexisting structure; indeed, that is all 'mutation' means, a change. If the reference sequence is to designate variants as mutants in this purely descriptive sense, it must be the original model from which imperfect copies arise. But this is obviously not true, whether we consider the Cambridge reference sequence or a replacement consensus sequence.

The mtDNA sequence of some arbitrary individual that is taken to be representative of a species or subspecies just in virtue of class membership readily becomes a "type" that characterizes that class. In one study, single African and Japanese sequences were selected, neither arbitrarily nor for their "averageness," but due to their presumed phylogenetic distance from each other and the European sequence (Horai et al. 1995). Reference sequences used to detect variation in sample sequences are often assumed to represent not just an "average" human, but a healthy one. Consensus sequences are believed to have functional significance; they are not idle statistical

---

[5] There is occasional, but relatively infrequent, recognition in research papers that the identification of mutations relative to the Cambridge reference sequence is one of convention. One paper specifically notes that two identified polymorphisms were "polymorphisms with respect to the published sequence of Anderson et al." (Holland et al. 1993, p. 549). In another, it is written: "Nucleotide changes are shown in the direction from the reference sequence (Anderson et al. 1981) to the sequences compared. The numbering system of Anderson et al. (1981) has been adopted" (Horai & Hayasaka 1990, p. 841). A third paper refers to "differences" (a bidirectional relation) rather than to "mutations" (a unidirectional relation): "Differences in nt sequence were identified in each of the cloned mtDNA fragments when sequences were compared between individuals, and with the published human mtDNA sequence (Anderson et al., 1981). All of these nt sequence differences consisted of single-nt substitutions" (Monnat & Reay 1986).

inventions. The aim of Marzuki and his colleagues, in proposing that the Cambridge sequence be replaced as a reference by a consensus sequence supplemented by a data base of normal variants, is to "provide a solid foundation for the definition of disease-related mutations in human mtDNA" (1991, p. 139). 'Normal' here does not mean average or typical, but healthy or functional. We find senses of the normal as unmutated original and the normal as healthy or functional to be conflated in one scientist's attempt to establish the authenticity of the reference sequence as an ancestral sequence. Ozawa (1995) argues that since it is of modern European origin, the Cambridge sequence cannot serve as the "normal standard mtDNA sequence" and suggests instead the sequence of mitochondrial Eve (mtEve), the African woman who is postulated to have lived about 200,000 years ago and to be the last common ancestor of modern humans.[6] Interestingly, Ozawa is motivated to find a better standard, not for population studies where it seems most appropriate, but as "the normal standard sequence for studies on mitochondrial diseases" (1995, p. 182). Hence, mtEve is assumed to represent a state of perfect health. From the original purity of the Garden of Eden has followed the disease and aging associated with mtDNA mutations.

2.3 Intersecting Notions of Normality

We have seen that the human mtDNA reference sequence is treated as "representative" in different ways in different contexts. The Cambridge reference sequence is sometimes considered to be a statistically normal or typical human mitochondrial genome that is likely to be found among members of a population. At other times, the Cambridge reference sequence is conceived to represent a standard of normal functioning or health against which the functional status of other human mitochondrial genomes can be assessed. The Cambridge reference sequence may even be used to represent a racial or ethnic "type." That these distinct notions of normality — statistical, functional, and

[6] Since Eve is no longer around to help us out, her mtDNA sequence is being estimated by sequencing the mtDNA of 48 individuals — Japanese, American-Irish, Australian-English, and Australian-Greek (Ozawa 1995).

typological — intersect in this way is not peculiar to the idea of a reference sequence as an expression of genetic normality. We find that statistical and functional notions of normality intersect and are conflated with one another in two other prevalent twentieth-century concepts of genetic normality: wild-type and consensus sequence. Yet, both inside and outside of biology, these are conceptually quite distinct senses of 'normal' that need not intersect: what works well may be infrequently found; racial or ethnic categorizations need not appeal to functional criteria; a "type" exemplar may rarely occur. I discuss these intersecting notions of normality within the context of the rise of probabilistic thinking in nineteenth-century biology.

2.3.1 Wild-Type Sequences and Consensus Sequences

As we saw with the mtDNA case study, a reference sequence, as a standard of comparison, can have more or less authority and more or less normative force. Since the Cambridge reference sequence became available over 15 years ago, there have been attempts to improve it — as a standard both of statistical and of functional normality. Two different conceptions of genetic normality are available to lend understanding to the notion of a reference sequence: wild-type and consensus sequence. "Wild-type" has long been the dominant conception of genetic normality, arising early in the twentieth century with the "Drosophilists." More recently, the techniques of "reverse genetics" have created a new conception of genetic normality, that of "consensus sequence." Despite these different origins, the concepts of wild-type and consensus sequence, like that of reference sequence, incorporate both statistical and functional senses of 'normal.' For example, 'wild-type' is defined by one author as "the predominant phenotype (appearance) in a population for any given trait" (Allen 1978, p. 149) and by another as "the original line of normally functioning individuals" (Judson 1996, p. 276).

The concept of wild-type arose in *Drosophila* genetics early this century. In the beginning, 'wild-type' really did mean wild. *Drosophila melanogaster* entered T. H. Morgan's lab at Columbia University in the fall of 1907 through Morgan's graduate student Fernandes Payne who had collected the flies, on Morgan's advice, by leaving ripe fruit out on the windowsill (Allen 1978, p. 147). These flies collected from the wild were then maintained in laboratory cultures. Early in 1910, a single white-eyed mutant male

appeared in one of Morgan's culture bottles (Morgan 1910, p. 149). In his 1910 paper, "Sex Limited Inheritance in Drosophila," Morgan contrasts the white-eyed "mutant" or "sport" with the "normal flies [that] have brilliant red eyes" (p. 120). Thus, wild-type organisms became understood as non-mutants. Morgan defines a reference in the paper to a "wild, red-eyed male" to mean "an individual of an unrelated stock" (p. 121). What Morgan no doubt was indicating by this is that the male to which he was crossing a white-eyed female descendent of his original white-eyed male was unrelated and would be of the "normal" type or pure stock and not harbouring a white-eye mutation. The results of his crossing experiments led Morgan to conclude that "wild" or "normal" males are heterozygous for red eyes and "wild" or "normal" females are homozygous for red eyes (pp. 121-122). He assumed there to be both genotypic and phenotypic uniformity in wild populations of *Drosophila* as well as in his wild-type stocks.

This assumption that *Drosophila* in the wild are genetically homogeneous was challenged by studies of natural populations carried out in Russia by Sergei S. Chetverikov and his students at the Kol'tsov Institute from 1922 to 1926 (Adams 1968). Laboratory geneticists, like Morgan, tended to think of the mutants that they maintained in culture bottles as abnormal variants that occur in nature but do not survive. Some naturalists, on the other hand, believed that mutant *Drosophila* strains were not merely abnormal but unnatural — that is, artifacts produced by artificial laboratory environments. Both these views were put to rest. Chetverikov's group captured *Drosophila melanogaster* and other *Drosophila* species in the wild and then brought them into the laboratory where they established inbred lines and found that a great deal of genetic variability underlies the appearance of phenotypic uniformity. In an important 1926 article, Chetverikov argues that the mutations observed in the laboratory also occur in nature but are hidden from view because most are recessives: "a species, like a sponge, soaks up heterozygous mutations while remaining ... externally (phenotypically) homogeneous" (in Adams 1968, p. 34). Should environmental conditions change, Chetverikov believed that the hidden variants may prove "decisive" for evolution (in Adams, p. 35). For Theodosius Dobzhansky, who drew upon Chetverikov's work in beginning his own studies of natural populations of *Drosophila pseudoobscura*, natural

populations ought to be regarded as genetically heterogeneous and "wild-type" flies ought to be conceived of as heterozygous at a large percentage of their loci and for a large number of different alleles.

Another assumption of classical geneticists was that there is a single wild-type allele at each locus and a variety of mutant forms. This view was adopted as early as 1913 when A. H. Sturtevant proposed the concept of multiple alleles — since several mutants seemed to occur at the same location on the linkage map, he believed that there might be an "original wild-type gene" that mutates to various forms (in Allen 1978, p. 181). The notation system that developed in Morgan's lab named genes for mutant, usually recessive, alleles and considered the wild-type to be "a standard of reference, usually symbolized as '+'" (Sturtevant 1965, p. 53). Biochemical and early molecular geneticists continued to use this notation in their work on microorganisms. Wild-type became understood explicitly in terms of a normal or original function. The "Drosophilists," although they recognized the inferior viability of the vast majority of their mutants, discerned and labelled mutant and wild-type organisms by phenotypic appearance. By contrast, wild-type *Neurospora* are able to grow on a minimal nutritive medium in the laboratory whereas mutants need a particular enzyme that they are unable to synthesize added to the medium. The abnormal state arises due either to spontaneous or radiation-induced mutation. However, the assumption that there is a single functional wild-type allele came into question as a result of protein electrophoretic studies of natural populations beginning in 1966. It became apparent that variant protein structure may in some cases have little or no effect on function. Due to the redundancy of the genetic code, even more variation at the level of genome might be expected to be functionally equivalent. Whether there is a single allele that might be called wild-type came into question.

Hence, many of the original assumptions about wild-type — for example, phenotypic uniformity, genotypic uniformity, original, functional norm, singular, "wild" or natural — are no longer credible. Yet, the term is used frequently: it appears 3112 times in a recent six month period (July 1997 to December 1997) in *Biological Abstracts*. 'Wild-type' in today's vernacular is understood to refer to an identified strain of a type of organism that has been standardized for research purposes. Robert E. Kohler (1994)

argues that experimental organisms are not merely biological creatures; they can also "be understood as technological artifacts that are constructed and embedded in complex material and social systems of production" (pp. 5-6). For example, "the 'standard' organisms — *Drosophila*, white mice and rats, maize, *E. coli* or *Neurospora* — ... have been reconstructed genetically through generations of selection and inbreeding into creatures whose genetic makeup and behavior are quite different from their natural ancestors" (p. 6). Ecological parameters such as activity and feeding are standardized to laboratory, not natural, environments. Natural variation is removed from laboratory strains of an experimental organism such as *Drosophila melanogaster* such that it is "domesticated" and becomes a "standard laboratory instrument" in order to standardize results. Preferable laboratory wild-type strains are not only genetically homogeneous but are highly canalized with their physiologies disrupted minimally by changes in environmental conditions. Nonetheless, for the laboratory geneticist, wild-type strains are still regarded as normal in both statistical and functional senses. Gene functions are discerned in experimental comparisons of wild-type with mutant or "knockout" organisms and laboratory results are presumed to be generalizable to nonlaboratory populations.

It is on the basis of normal (whether statistical or functional) organismal phenotypes that genotypes, genes, and nucleotide sequences are identified as wild-type. For example, in clinical studies, normal and mutant genes are distinguished from one another depending on whether they are found in healthy or diseased individuals. Consensus sequences, on the other hand, are identified from the bottom up. 'Consensus sequence' arrived on the scene in the late 1970s when DNA sequence data first became available. Despite its youth, the term is used in several different ways. Standard Oxford dictionary definitions of 'consensus' are helpful in trying to sort these out. The first definition given is: "General concord of different organs of the body in effecting a given purpose." This functional sense of 'consensus' corresponds with the meaning of 'consensus sequence' as it first appears in the late 1970s and early 1980s. The focus is on the structure-function relationships of various genomic elements. What researchers are seeking to identify is particular DNA sequences that have characteristic functions and that are shared by different species (are homologous) and/or occur in different parts of the genome. Structural similarities are identified in areas of the genome that are expected to

share certain functions. For example, consensus sequences are identified for RNA splice sites (Rogers and Wall 1980), chromosomal "hotspots" for transposon insertion (Halling and Kleckner 1982), progesterone-receptor binding sites (Mulvihill et al. 1982), promoter regions in *E. coli* (Navre and Schachman 1983), the TATA box in eukaryotes (Weaver et al. 1982), translational start sites in mRNA (Kozak 1986), and glucocorticoid regulation of gene expression (Hardman et al. 1984). The structural similarities that characterize particular consensus sequences are termed "sequence motifs." The Oxford dictionary defines 'motif' as "a distinctive feature or element of a design or composition in art and literature." Hence, the idea of functional design is integral to the concept of a consensus sequence. Consensus sequences are also studied experimentally to see if they are necessary and/or sufficient for the particular function — for example, heat-induced transcription — to be carried out. Altered functions are understood in terms of deviations from the "perfect" consensus sequence. For example, there is closer conformity to a six base-pair consensus sequence for transposon insertion at "hot spots" than at other insertion sites and the sequence tends to be absent from areas of the genome where transposons do not insert. In all of these cases, it is expected that functional elements will be subject to similar structural constraints and, therefore, that these structural elements will occur frequently — that is, the functional and statistical notions of normality will coincide.

The second definition the Oxford dictionary lists for 'consensus' is: "Agreement in opinion." This statistical sense of 'consensus' corresponds to a second way in which the term 'consensus sequence' occurs in biology. Consensus sequences are statistical averages: a consensus sequence includes at each nucleotide position the nucleotide base that occurs most frequently at that position when multiple sequences are compared. Sometimes these sequences derive from the same individual's DNA. As we saw in section 2.1.3, the "agreement of opinion" here involves ensuring that technical errors are not made in the large-scale sequencing of DNA by sequencing overlapping DNA clones from the same individual. The sense of 'consensus sequence' we encountered in the mtDNA case study refers instead to the production of a sequence based on the comparison of DNA samples taken from different individuals. Here, the consensus sequence represents a statistical average; it is, in a sense, an "agreement in opinion" reached by a democratic sampling of individuals. The representation may be purely statistical. In

a recent patent application for the "normal" BRCA1 gene, one genetic testing company argued for the superiority of its proposed BRCA1 consensus sequence (based on a comparison of the gene's sequence in five "normal" individuals without family histories of breast or ovarian cancer) over the wild-type sequence offered in a rival company's application because it is "the most likely *BRCA1* sequence to be found in the majority of the normal population" (Marshall 1997a, p. 1874). Consensus sequences have also been used to define mutations as deviations from the statistical average. For example, a consensus sequence was derived from six of a "family" of repetitive DNA sequences, called R sequences, with an estimated 100,000 of these distributed throughout the haploid mouse genome with *possible* functional roles: "The individual R sequences have an average divergence from the consensus sequence of 12.5%, which is largely due to point mutations and, among those, to transitions" (Gebhard et al. 1982, p. 453). "Tentative human consensus sequences" (THCs) are being compiled for gene transcripts by sequencing portions of cDNAs that are produced by reverse transcription from mRNA samples that originate in diverse individuals. THCs are contigs that are assembled from ESTs to approach full-length transcripts of expressed genes using a computer algorithm that ensures that only ESTs that meet "stringent overlap criteria" are included (Adams et al. 1995, p. 7). Sometimes, considerations of function are explicit. A consensus sequence might identify positions where nucleotide substitutions can occur without loss of function (Zyskind et al. 1983). Or, an already-identified consensus sequence might be evaluated to see if it is the "optimal" sequence from a functional standpoint by determining what happens when individual base substitutions are made (Kozak 1986).

Hence, like "reference sequence," "wild-type" and "consensus sequence" are concepts of genetic normality that convey both statistical and functional notions of normality and often conflate these.

2.3.2 Conceptions of Normality in Nineteenth-Century Biology

Although statistical, functional, and even typological notions of normality intersect and become conflated with one another in concepts of genetic normality such as reference sequence, wild-type, and consensus sequence, the distinctions between these senses were central to developments in nineteenth-century biology. It is Adolphe Quetelet, astronomer

turned "social physicist," who is credited with initiating the probabilistic revolution in the sciences by introducing the statistical method to the social sciences early in the nineteenth century. In this, he was opposed by philosopher Auguste Comte who held that the social realm, no less than physics, astronomy, chemistry, and physiology, requires predictive laws based on observations — an impossibility if observations arise due to chance (Cohen 1987). Similarly, Claude Bernard rejected the validity of statistical reasoning in physiology because he sought to establish physiology as a causal science for which an underlying determinism must be assumed (Coleman 1987). Francis Galton's efforts to introduce statistical thinking to biology were stimulated by Quetelet's use of the error law to study human variation (Porter 1986, pp. 135ff).

Bernard was vehemently opposed to the use of statistics in biology and medicine. As he wrote in *An Introduction to the Study of Experimental Medicine*, "scientific law can be based only on certainty, on absolute determinism, not on probability" (1957, p. 136).[7] Statistical methods yield "conjectural," and not "true" or "sure," sciences (p. 139). Biological science exists only because "[a]bsolute determinism exists ... in every vital phenomenon" (p. 65). Bernard viewed organisms as "living machines": "a living organism is nothing but a wonderful machine endowed with the most marvellous properties and set going by means of the most complex and delicate mechanism" (p. 63). Experimental study analyzes the organism as one takes apart a machine in order to discover the conditions or "hidden springs" that are necessary for a given phenomenon to occur. In the same way in which inanimate machines are found to function normally or to malfunction, organisms exist in either healthy or diseased states: "in nature there can be only order and disorder, harmony or discord" (ibid.). The laws of mechanics apply to inanimate machines whether these are or are not working properly; similarly, whether organisms are healthy or diseased, their physiological processes are governed by the same "vital" laws: "Since all these phenomena [physiological, pathological, and therapeutic] depend on laws peculiar to living matter, they are identical in essence and vary only with the various conditions in which phenomena appear" (p. 193). Altered conditions,

---

[7] The original French version of Bernard's text was published in 1865 and first translated into English in 1927.

especially with respect to organisms' internal environments, explain the manifestations of health and disease: "By normal activity of its organic units, life exhibits a state of health; by abnormal manifestation of the same units, diseases are characterized" (p. 65). Diseases occur where the normal regulatory mechanisms of the internal environment have been disrupted and the harmonious interdependence of parts is lost. Bernard also acknowledged that nonpathological physiological differences could arise in members of the same species and race due to differences in internal environments; he referred to these as individual "predispositions" or "idiosyncracies." Statistical averages, Bernard believed, served to obliterate these differences and to discourage their scientific (causal) explanation.

In contrast to Bernard, Quetelet and Galton were concerned with properties of individuals only insofar as these individuals are conceived to be members of a population or race. It was as an astronomer that Quetelet was first introduced to the theory of probability in the 1920s by Laplace who encouraged him to apply statistical methods to the study of society (Diamond 1969, p. viii). Quetelet began this analysis by tabulating averages of physical measures (for example, height or weight) or frequencies of social occurrences (for example, marriage or criminal behaviour) and determining their relationships to various parameters. Upon finding these relationships to be stable from one year to another and one country to another, Quetelet became convinced that he had discovered "laws" that could sustain the study of "social physics" (Lécuyer 1987). Hence, from the beginning, Quetelet assumed an underlying determinism. He believed that laws operate at the level of society that are analogous to those that govern the solar system although individuals, like planets, may be disturbed in their movements by "perturbing forces" (Diamond 1969, p. viii). In his (1835) *Sur L'Homme, et le Développement de ses Facultés*, Quetelet introduced his conception of "*l'homme moyen*" or the "average man." In *Sur L'Homme*, Quetelet initially presents "*l'homme moyen*" as a statistical abstraction — a pretend-individual who instantiates the properties of an entire population. This allowed him to ignore the peculiarities of actual individuals and to focus on generalities that emerge when many individuals are studied:

> The social man ... resembles the centre of gravity in bodies: he is the centre around which oscillate the social elements — in fact, so to speak, he is a fictitious

> being, for whom every thing proceeds conformably to the medium results obtained for society in general. It is this being whom we must consider in establishing the basis of social physics, throwing out of view peculiar or anomalous cases. (Quetelet 1842, p. 8)[8]

In this way, Quetelet's "*l'homme moyen*" is described as a "statistical composite of the physical, moral, and intellectual traits of the entire society" (Daston 1987, p. 303) and as the fictive recipient of dispositions or numerical propensities (called "penchants" and "tendencies" by Quetelet) that cannot belong to actual individuals (Krüger 1987, p. 74). "*L'homme moyen*" could be no more than a mathematical abstraction insofar as it is based on properties that belong properly to populations and not to individuals.

However, even in 1835, Quetelet's "*l'homme moyen*" represented more than a descriptive device: when considered abstractly, Quetelet believed his "average man" to exhibit some "remarkable properties" (1842, p. 96). "*L'homme moyen*" represented, for Quetelet, not just a statistical average but a type that is characteristic of a given race or nation, or humankind generally: "Every race has its peculiar constitution, which differs from this [human type] more or less, and which is determined by the influence of the climate, and the habits which characterize the average man of that particular country" (p. 99). "*L'homme moyen*" also represented a standard of physical health for Quetelet:

> if the average man were completely determined, we might ... consider him as the type of perfection; and everything differing from his proportions or condition, would constitute deformity and disease; everything found dissimilar, not only as regarded proportion and form, but as exceeding the observed limits, would constitute a monstrosity. (ibid.)

Quetelet recognized that physicians' reliance on such a standard would inevitably introduce error because "general laws referring to masses are essentially imperfect when applied to individuals" (ibid.). However, he also believed such comparisons would be helpful in most cases and that physicians, who usually do not see their patients except when they are sick, have no other basis for making clinical judgements. "*L'homme moyen*" served as a moral and intellectual, as well as physical, ideal — not absolutely, because human nature progresses, but for a given time and place: "in the circumstances in which he is found, [the average man] should be considered as the type of all which is

---

[8] 1842 is the date of the first English translation of Quetelet's *Sur L'Homme*.

beautiful — of all which is good" (p. 100). Quetelet explained genius in terms of the degree to which literary, scientific, and artistic greats approximate "*l'homme moyen*" — "great men" are the "best representatives" of an age (p. 101). Quetelet recognized that his "average man" exists neither as a statistical abstraction nor as an ideal type:

> an individual who should comprise in himself (in his own person), at a given period, all the qualities of the average man, would at the same time represent all which is grand, beautiful, and excellent. But such an identity can scarcely be realised, and it is rarely granted to individual men to resemble this type of perfection, except in greater or less number of points. (p. 100)

However, the "is" and "ought" had become entangled: individual peculiarities are not just statistically rare but deviations from an ideal physical, moral, and intellectual racial type.

It is argued that, after 1840, Quetelet began to focus less on a trait's mean value in a population and more on its distribution (Lécuyer 1987). In 1843, he advanced the theory that all human traits are distributed according to the "law of accidental causes" (Diamond 1969, p. xi). Ian Hacking (1990) argues that, at this time, "*l'homme moyen*" was transformed from an abstract property of a population that expresses various statistical regularities to a real property that is produced by genuine causes — a natural kind, in other words. "*L'homme moyen*" became "*l'homme type*," ideal not as a statistical abstraction or "golden mean" but as nature's essential type. As each copy of a statue is imperfect, so to is "every real man ... an imperfect replicate" of "*l'homme type*" (Porter 1986, pp. 106-108). This development in Quetelet's thought amounted to a reversal of the way in which astronomers construed the observational "law of errors." Astronomers understood the normal curve to reflect repeated measurements of a constant value (the position of a planet, for example) confounded by observational errors. Quetelet attributed mean values to the operation of constant causes and deviations from the mean to perturbations of constant causes by accidental causes which compensate for one another in direction and degree over the long run. In other words, Quetelet lent an ontological interpretation to what had been an epistemological one (Sober 1980, p. 365):

> it was as if nature had aimed at an ideal value but only obtained somewhat disturbed results. The carrier of the ideal values could then be interpreted as the type nature had aimed at. In this case the man (or mean man) turned out to be "l'homme type." The distribution appeared as a law of nature that serves to preserve the species. (Krüger 1987, p. 75)

Insofar as nature operates to preserve the mean and the distribution about the mean, human variability is itself understood to be law-like and "natural" although produced by "accidental causes" (Krüger 1987). In Quetelet's words: nature "confers an infinite variety to everything that breathes, without impairing its principles of preservation" (in Krüger, p. 76). That deviations fall within a regular pattern of distribution shows that they arise from interferences with natural laws but do not lie outside these laws (Porter 1986, p. 100). With additional knowledge, it would be possible to explain each deviate value in terms of these "accidental" causes (ibid., p. 105).

The error curve is arrived at in two different ways. The first route, the astronomers' "law of errors" taken by Quetelet, has already been noted. The second route is through the binomial distribution of values that is associated with repetitive coin tosses. On the binomial interpretation, population data are normally distributed because multiple interacting independent causes are responsible for the development of traits in individuals. This seems to be Galton's interpretation.[9] Hacking (1990) argues that, with Galton, statistical laws became autonomous. By 'autonomous,' Hacking means that although the laws may, in principle, be reducible to underlying causes, such a reduction is unnecessary for them to be explanatory (pp. 181-182). What Hacking believes to be Galton's original contribution is his insight that the distribution of a trait in a population can be explained in terms of its distribution in the population in preceding generations without a concomitant (deterministic) need to appeal to underlying causes (p. 186). However, as Galton worked out these statistical laws, he was also formulating a mechanistic model of heredity that helped to explain the population-level parameters.[10]

In his (1869) *Hereditary Genius*, Galton offered support for his cousin Charles Darwin's identification of hereditary particles called gemmules the previous year. Gemmules are transmitted from generation to generation; some of these will be "patent" and some of these will be "latent" in given individuals. Galton stressed that "the theory

---

[9] Hacking (1990) however argues that Quetelet interpreted his "discovery" that all human traits are normally distributed in terms of the binomial distribution, a movement resisted by Galton.

[10] Hacking does note that Galton's treatment of statistical laws of heredity as autonomous does not mean that he "gave up his belief in some underlying determinism, nor even that he gave up the model of petty independent influences" (1990, pp. 185-186).

of Pangenesis brings all the influences that bear on heredity into a form, that is appropriate for the grasp of mathematical analysis" (p. 373). Pangenesis makes sense of the "stability of types" and the transmission of variation from parents to offspring. Individuality is understood as "a segregation of what already existed [in the "stock" of nature], under a new shape, and as a regular consequence of previous conditions" (p. 376). For Galton, gemmules, like balls in a urn, account for the regularity of the normal distribution curve as they are inherited from one generation to the next:

> It remains that I should say a very few words on the principle of the law of deviation from an average, or, as it is commonly called, the law of Errors of Observations, due to La Place. Every variable event depends on a number of variable causes, and each of these, owing to the very fact of its variability, depends upon other variables, and so on.... Also, by the very fact of each of these causes being a variable event, it has a mean value, and, therefore, it is ... an even chance in any case, that the event should be greater or less than the mean. Now, it is asserted to be a matter of secondary moment to busy ourselves in respect to these minute causes, further than as to the probability of their exceeding or falling short of their several mean values, and the chance of a larger or smaller number of them doing so, in any given case, resembles the chance, well known to calculators, of the results that would be met with when making a draw out of an urn containing an equal number of black and white balls in enormous numbers. (p. 382)

Galton later rejected elements of Darwin's theory of pangenesis, adopting a theory of "hard heredity" in which the germinal material is confined to the reproductive organs (Porter 1986) and "transmitted unchanged from one generation to the next" (Bowler 1989, p. 64). He further elaborated his theory of hereditary transmission in 1873. Statistical sampling of gemmules happens twice. First, of the gemmules transmitted from parents to offspring, only a small percentage become the genetic material or "stirp" that determines the development of the individual; the remaining gemmules remain latent. Second, only a percentage of the total gemmules — both active and latent — are passed from each parent to the offspring and the remainder perish (see Porter 1986, pp. 283-4).

Various commentators note that whereas Quetelet had focused mostly on mean values, Galton was interested in the entire distribution of values. With Galton, variation within a population became real and legitimate: "To most persons Variability implies something indefinite and capricious. They require to be taught that it, like Proteus in the old fable, can be seized, securely bound, and utilized; that it can be defined and

measured" (1907, p. 16). Variation was no longer, as it was for Quetelet, "error" or deviation from an optimal type. Since the constancy of the normal distribution of a trait in a population from one generation to the next is indicative of its hereditary basis, there is a sense in which the entire bell-shaped distribution of values is "normal." Galton did, however, believe that means define specific types of individuals — such as criminal or consumptive types — as well as racial types (Porter 1986, pp. 139-140). For example, he held that "the average ability of the Athenian race is, on the lowest possible estimate, very nearly two grades higher than our own — that is, about as much as our [Anglo-Saxon] race is above that of the African negro" (1869, p. 342). "Ideal mean types," in this way, characterize families and races and denote their relative worth:

> The processes concerned in simple descent are those of Family Variability and Reversion.... By family variability is meant the departure of the children of the same or similarly descended families from the ideal mean type of all of them. Reversion is the tendency of that ideal mean type to depart from the parental type, "reverting" towards what may be roughly and perhaps fairly described as the average ancestral type. (1877, p. 513)

"Ideal mean types" do not represent, though, as they did for Quetelet, ideal values from which members of the characterized population deviate. Whereas Quetelet had focused on how social forces might be manipulated in order to decrease deviation about the mean, Galton was interested in the "exceptional man" and how heredity might be manipulated to increase the relative frequencies with which "exceptional" traits such as high intelligence would appear in the population.[11] In *Hereditary Genius*, Galton adopted Quetelet's use of the error law specifically to demonstrate how the rarity of exceptionally talented individuals and the prevalence of mediocre individuals is to be expected (Porter 1986, p. 142).

This brief excursion into the nineteenth century demonstrates that statistical, functional, and typological notions of normality are conceptually and historically quite distinct from one another. Bernard's view of the organism as "living machine" attributes to it goals that it either manages or fails to accomplish. Although machines continue to

---

[11] It should be noted that, in 1848, Quetelet further distinguished between "stationary" (physical) and "progressive" (mental) qualities in order to recognize that above average intellect, for example, ought not be regarded as a defect (Diamond 1969, p. xii).

work when a functionally equivalent part is substituted for another, if they are well-designed, most substitutions of parts will be damaging. Hence, on Bernard's strictly deterministic conception of normality as proper functioning at the level of individuals, variation is invariably, at least potentially, harmful deviation. We will see in Chapter Four that functional notions of normality continue to dominate laboratory sciences such as physiology and molecular genetics which aim to discover the universal laws and causal mechanisms that govern the behaviours of individual organisms and their component parts.

Statistical notions of normality concern the distributions of traits in populations and not their presence or absence in individuals. If the normal distribution of values is given a wholly statistical interpretation, it is possible to treat the entire bell-shaped curve itself as normal. Although values that lie close to the tails of the distribution occur far less often than values that lie close to the mean, they are abnormal only in that they are infrequent; they are not deviations from anything other than a statistical norm. This represents the trend that was initiated but not completed by Galton. With contributions between 1902 and 1918 from G. Udney Yule, H. Nilsson-Ehle, Edward M. East, and R. A. Fisher, it was finally recognized that the approaches of the biometricians (Galton and his successors) and the Mendelians were compatible.[12] Continuously varying traits in a population can be explained in terms of multiple Mendelian factors segregating in a population. This is how quantitative geneticists understand continuous variation today, with nongenetic factors recognized to contribute to the distribution as well. Variation is simply difference. There is no *a priori* expectation that unlikely or unusual values require explanation. In a sense, they are "certain" to occur just as, given an adequate number of trials in which a coin is tossed ten times, there is "certain" to be an occasion upon which all ten tosses come out as heads. On the statistical view, the definition of normal and abnormal based on numbers of standard deviations from the mean is recognized to be purely arbitrary and of no necessary functional significance.

On Bernard's functional conception of biological normality, virtually all variation between individuals is deviation. Given a purely statistical interpretation of the normal

---

[12] See Provine (1971) for a complete history of this period.

distribution of traits, for which Galton's thought was transitional, variation in a population is simply difference. Quetelet's approach falls midway: variation is harmful deviation, but deviation that is expected because deterministic forces operate at the level of the population, not the individual. Once the normal distribution curve is interpreted in entirely probabilistic terms, the illegitimacy of Quetelet's concepts of "*l'homme moyen*" and "*l'homme type*" is fully appreciated. Yet, these concepts were criticized even in Quetelet's time. Cournot objected to the statistical conception of "*l'homme moyen*" on the grounds that the mean values of all traits would be incompatible if combined (Lécuyer 1987). Similarly, it was argued that "*l'homme moyen*" is a "mathematical fiction" from which it is impossible to infer anything about the properties of actual individuals (Porter 1987). Bertillon argued that "*l'homme moyen*" would represent mediocrity, a "*type de la vulgarité*," and not a moral or intellectual ideal (Lécuyer 1987, p. 330). Associated with the rejection of Quetelet's views in Germany was a challenge that Theodore M. Porter (1987) portrays as centering on the question of ontological priority. Quetelet's German critics argued that variation ought not to be considered as "mere error." It is because society is composed of heterogeneous individuals that statistics are necessary in social science. Where these heterogeneous individuals are similar, regularities emerge at the higher level. This runs counter to Quetelet's assumption that individuals are alike because they are acted on by constant social forces and differ only by accident. Despite these long-standing critiques of Quetelet's concept of "*l'homme moyen*," we find in molecular biology's recent concept of consensus sequence a similar notion of a statistical average that is in some way authoritative for the population.

2.4 Summary

From the mtDNA case study, it is clear that when the complete nucleotide sequence of "the" human genome is obtained early next century, researchers in different fields of biology will make use of the available reference sequence in various ways. Its representational meaning will be neither univocal nor fixed. Meaning will accrue to the

reference sequence as it is used, according to the ways in which it is used. Certainly, evidence is strong that the reference sequence, likely with revisions over time, will serve as a normative standard. Insofar as functional and statistical notions of normality intersect in concepts of genetic normality such as wild-type and consensus sequence, the reference sequence is likely to be considered normal in multiple ways as well. Slippage from descriptive to evaluative senses — from what is usual, to what works, to what is desirable — all too easily occurs: an arbitrary sequence becomes the essential sequence; variation is viewed as deviation, not simply as difference; mutations are understood not as changes but as structural damage.

Part of the confusion may lie in the fact that words like 'normal, 'representative,' 'standard,' and 'reference' are "waffle words" in everyday language as well as in biology. These terms are similar in that they are used, often synonymously, both to describe and to evaluate. However, the intersection and conflation of descriptive and normative notions of genetic normality, genetic variation, and genetic mutation cannot, for the most part, be attributed to ignorance or semantic confusion. As we will see in Chapters Three and Four, the reasons that the "is" and the "ought" intersect in human molecular genetics are basic to the conceptual foundations of the discipline itself. It would therefore be unreasonable to assert that the conflation of the "is" and the "ought" lies with the failure of biologists to be cognizant of the conceptual and historical discontinuities between statistical, functional, and typological notions of normality that date to the nineteenth century. Nor can it be ruled by decree that only statistical norms are scientifically objective or that only functional norms are biologically meaningful. This would centre on extrascientific concerns about what counts as science and what counts as biology's proper object of study. Rather than any attempt to sweep away blindly the conceptual cobwebs in which biological understandings of human genetic variation are entangled, the necessary foray is one that traces along the fibres of these webs to try to discover the sources of their entanglements.

This path leads to the evolutionary and clinical contexts of research in human molecular genetics. Chapters Three and Four focus on whether genetic variation is to be understood as value-neutral statistical difference, deviation from a biological (functional) norm, or deviance from nonbiological aesthetic, moral, social, or cultural norms. As

Dobzhansky's widely-quoted saying attests: "Nothing in biology makes sense except in the light of evolution" (1973). Evolutionary theory legitimizes the conflation of what is usual with what works because it assumes that what is usual has worked in the past and was therefore favoured by natural selection. Chapter Three, "The Evolutionary Context: Is Genetic Variation Difference or Deviation?," by situating the concept of a DNA reference sequence within a dynamic evolutionary context, recognizes its contingency as a norm. Chapter Four, "The Clinical Context: Is Genetic Variation Deviation or Deviance?," situates the concept of a DNA reference sequence within its clinical and cultural contexts. Insofar as clinical judgements of health and disease incorporate nonbiological aesthetic, moral, social, and cultural values, any genetic standard of normal functioning and health that follows from such judgements is not just evolutionarily but also culturally contingent.

Chapter Three

The Evolutionary Context: Is Genetic Variation Difference or Deviation?

> The essential quality of life is living; the essential quality of living is change; change is evolution: and we are part of it. (Wyndham 1958, p. 196)
>
> a mature physicist, acquainting himself for the first time with the problems of biology, is puzzled by the circumstance that there are no 'absolute phenomena' in biology. Everything is time-bound and space-bound. The animal or plant or micro-organism he is working with is but a link in an evolutionary chain of changing forms, none of which has any permanent validity. (Delbrück in Mayr 1961, p. 1502)
>
> To the extent that living beings diverge from the specific type, are they abnormal in that they endanger the specific form or are they inventors on the road to new forms? One looks at a living being having some new characteristic with a different eye depending on whether one is a fixist [*fixiste*] or a transformist. (Canguilhem 1989, p. 141)
>
> But life is change, that is how it differs from the rocks, change is its very nature. Who, then, were the recent lords of creation, that they should expect to remain unchanged? (Wyndham 1958, p. 182)

In Chapter Two, we saw that it is possible to distinguish several logically distinct notions of normality: what is frequent, original, functional, or ideal. These different senses of 'normal' are often conflated, in biology as well as in everyday parlance. This was illustrated in the last chapter's look at studies which compare sampled mitochondrial DNA (mtDNA) sequences to the human mtDNA reference sequence. However, in biology, this only sometimes amounts to semantic confusion. The phenotypes encountered most frequently tend to be those that have been most successful in evolution through natural selection. And if an organism is well-adapted to its environment, random departures from this state are unlikely to represent improvements to function. What is frequent and what is "original" are, for the most part, what functions optimally. Adaptive evolution, through natural selection among available variants, often provides the justification for what in nonbiotic universes of discourse might constitute poorly chosen language. This chapter places the Human Genome Project's goal to establish a human DNA reference sequence in its evolutionary context — specifically, within the context of two important

controversies in twentieth-century evolutionary genetics: the classical-balance and neutralist-selectionist debates.

Despite the fact that adaptive evolution justifies the intersection of several possible senses of 'genetic normality,' evolutionary biologists and philosophers of an evolutionary bent have been critical of the HGP's aim to obtain the complete sequence of "*the* human genome." The implicit assumption that there exists some ideal "species-type" genome instantiated, if imperfectly, by particular individuals is a throwback, they say, to pre-Darwinian, Platonic, essentialist, and typological thinking. This thinking sees mutations as abnormal changes in genetic structure and not simply as changes. Such changes are abnormal, not only in that they are rare, but because they represent "errors" in the genetic code or "damage" to the proper genetic structure. Hence, the genetic variation within a species or subpopulation, the ultimate source of which is mutation, is not simply difference, but deviation from some adaptive norm. What this forgets, Camille Limoges writes, is that "[g]enetic variation is the source of evolution.... it is genetic 'errors' that made us as a biological species" (1994, p. 124). David Hull, in a similar vein, argues that typological thinking ignores the fact that "[t]he essence of a species is to have no essence" (1994, p. 215). That the HGP's composite genome has been named "Linnaeus" after the 18th century systematist who believed until close to the end of his life in the immutability and divine creation of species does little to deflect such criticism (ibid.).

Although the HGP's mandate is unabashedly molecular, James Griesemer (1994) stresses the need to resolve these interdisciplinary theoretical differences; after all, molecular and evolutionary geneticists deal with the same organisms/genomes/alleles and have interdependent explanatory frameworks. In this chapter, I take a different tack. I argue that the apparent gap between the molecular and evolutionary approaches can best be understood in terms of molecular biology's assumption of a particular set of evolutionary beliefs, rather than its ignoring of evolutionary considerations altogether. For example, it makes perfect sense to speak of alleles or genomes as "normal" or "deviate" and to underscore the harmfulness of mutations or "genetic load" if one believes that adaptation, not chance, is the predominant mechanism of evolutionary change and that natural selection acts primarily to eliminate, rather that to preserve, variation within populations. Bets are off, however, once these assumptions are contested. And contested

they are. I agree with Griesemer that it is desirable to aim to achieve theoretical consistency within biology. However, it is misleading to oppose evolutionary and molecular genetics, as Griesemer, Hull, Limoges and others do, as if evolutionary genetics is itself monolithic and in possession of a unified theoretical structure.

Ever since R. A. Fisher, J. B. S. Haldane, and Sewall Wright laid the foundations for the discipline early this century, theoretical population geneticists have failed to achieve a consensus on some key conceptual issues. Three important debates characterize, and have shaped the course of, twentieth-century developments in evolutionary genetics. These are the drift-selection debate associated with Wright and Fisher, the classical-balance debate between H. J. Muller and Theodosius Dobzhansky, and the still-ongoing neutralist-selectionist debate. In this chapter, I focus on the latter two of these three controversies, and refer to the drift-selectionist debate only to clarify its historical and conceptual continuities with the others. The chapter is divided into three sections. First, I visit the classical-balance debate which centred on many of the same issues implicated in today's evolutionary criticisms of the HGP: Is there a "normal" genome? Are mutations "bad"? Is variation deviation? Second, I look at the current neutralist-selectionist debate and examine its conceptual continuities with its classical-balance predecessor. I focus on issues that remain unresolved in our understandings of evolutionary meanings of intraspecific genetic variation. Third, I summarize the evolutionary beliefs that are consistent with the HGP's approach to genetic variation and consider reasons that might account for molecular genetics' subscription to such beliefs. I evaluate evolutionary criticisms of the HGP's goals to map and to sequence "the" human genome as a throwback to pre-Darwinian, Platonic, essentialist, and typological thinking in view of Dobzhansky's similar criticisms of Muller.

3.1 The Classical-Balance Debate

The rather acrimonious Muller-Dobzhansky debate was initiated by Muller's 1949 presidential address to the American Society of Human Genetics entitled "Our

Mutations"[1] and persisted until his death in 1967. The labels "classical" and "balance" follow Dobzhansky's characterization of the two opposing positions in his Cold Spring Harbor address of 1955.[2] My interest in returning to the scene of this controversy lies in the similarity of Muller's conceptions of a "normal" genome, and the harm posed by mutation and genetic variation, to those of today's molecular biologists. Reminiscent of Hull's and Limoges' criticisms of the HGP, Dobzhansky had accused Muller of pre-Darwinian, Platonic, essentialist, and typological thinking. I begin this section by providing an outline of the terms of the classical-balance debate. I then examine more specifically Muller's and Dobzhansky's differing views on "genetic normality," the harmful effects of mutations, and the adaptive value of intraspecific variation.

3.1.1 The Terms of the Debate

From the outset, it is important to recognize that Muller and Dobzhansky shared a common theoretical framework in that both were committed neo-Darwinists. As neo-Darwinists, they held evolution to be a slow, gradual process, the result of the action of natural selection on the variation in quantitative traits that is furnished by randomly occurring mutations and, in sexually reproducing species, recombination. Like other geneticists of their time, they ignored the possibility that genetic mutation could occur without any impact on phenotype. They believed that phenotypic variation, even where it appears to be of no adaptive significance, is unlikely to be neutral with respect to selection. Although both supported a role for drift in determining the genetic composition of small isolated populations, on balance, Muller, and Dobzhansky by this time, were proponents of strongly adaptationist views of evolution.[3] Organisms evolve as a result of the incorporation by their genomes of the very infrequent mutations that prove beneficial, most often "under rare conditions or in rare combinations with other

---

[1] Subsequently published as Muller (1950b).

[2] Subsequently published as Dobzhansky (1955a).

[3] Although Dobzhansky initially favoured a predominant role for random drift in evolution (and, indeed, popularized Wright's shifting balance theory), he played a major role in what Gould (1983) refers to as the "hardening of the evolutionary synthesis" and came to attribute increased relative importance to adaptation.

mutations" (Muller 1950a, p. 174) as might happen in the event of environmental change. Although Muller and Dobzhansky were both adaptationists, they emphasized different dynamics of selection. It must be stressed that their disagreements centred on the relative weight that each was prepared to attach to a particular mode of selection, and not whether the mode is at all operative.[4] Holding adaptive evolution to be the result of the incorporation of very infrequently occurring beneficial mutations into the genome of a species especially should the environment change, Muller and Dobzhansky agreed that positive or directional selection is "the most important agency in bringing about long-term evolutionary changes" (Dobzhansky 1962, p. 156). Where they disagreed was over the relative importance of the more prevalent nondirectional forms of selection.

Muller believed that natural selection is predominantly negative or "purifying" and acts to preserve the adaptive norm by eliminating genetic variation in a population. This type of selection is referred to by I. I. Schmalhausen (1949) as "stabilizing" (as opposed to "dynamic") and by C. H. Waddington (1957) as "normalizing." Muller defends this view especially in his 1950 article, "Evidence of the Precision of Genetic Adaptation,"[5] where he argues that organismal traits are well-adapted for their circumstances not only in type, but in degree. That almost complete saturation of normal alleles and dosage compensation for sex-linked traits in *Drosophila* have evolved shows that even very small phenotypic departures from the norm that are imperceptible to us must detract from the organism's fitness.[6] Natural selection gradually "whittles" away at these slight, but evidently disadvantageous differences, to maintain the trait at its optimal value. As a

---

[4] Beatty (1987) notes that this is a feature that characterizes most disputes in biology: What is contentious "is not whether nature *always* follows this course or that, but rather the *relative importance* of the various courses that nature follows" (p. 293).

[5] The lecture, as part of the Harvey Lecture Series, was delivered on February 19, 1948.

[6] Normal alleles are observed to operate at close to full saturation in that in most cases normal genes are dominant to their mutants and there is little difference between the homozygote with two normal alleles and the heterozygote with only one. Muller surmised that this functions to protect the normal characteristic from variability due to differences in environmental and genetic backgrounds. This goes a long way to compensate the male for having only one copy of any allele carried on the X-chromosome, and yet a system of dosage compensation that permits sex-linked traits to be expressed to a similar degree in both male and female has also evolved, demonstrating once again, the adaptive significance of even "subliminal" differences.

result, stabilizing selection maintains structures that would otherwise degenerate due to "mutation pressure": "natural selection [is] a process which not only leads, sometimes, to further adaptations but which is everywhere actively at work in maintaining all things biological that merely continue in existence" (1973a, p. 190).

Dobzhansky agreed that stabilizing selection is an important negative evolutionary force that protects the adaptive norm by eliminating poorly adapted mutants. However, unlike Muller, he considered balancing selection to be the more important evolutionary force, at least for sexual outbreeding organisms. Whereas "purifying" or stabilizing selection operates to eliminate genetic variation in populations, balancing selection creates stable polymorphisms that guarantee its maintenance. Balancing selection is of two types: heterotic and nonheterotic. Dobzhansky tended to emphasize heterosis but admitted that the relative proportions of each form of balancing selection remained an open question. Heterotic balancing selection operates where the fitness of heterozygotes is superior to that of homozygotes for any of the relevant alleles. Such loci are referred to as overdominant. A familiar example in humans is the superior fitness of carriers of the allele for sickle-cell anemia in malarial environments. Dobzhansky interpreted experimental data from *Drosophila* to show that the superior fitness of heterozygotes lies in their "versatility," or ability to "live well in a wider variety of environments." The "normal" homozygote demonstrates superior fitness only within a narrow range of environments. Frequency-dependent selection is one form of non-heterotic balancing selection: an example is the selection of mimetic polymorphisms in butterflies. Nonheterotic balanced polymorphisms can also be generated by the fluctuating selection coefficients associated with seasonal changes or diverse habitats; Dobzhansky referred to this as "diversifying" selection (1962, p. 288).

Muller accepted that balancing selection maintains polymorphisms at a small number of loci. In humans, he held short-sightedness to be an example of a stable frequency-dependent polymorphism like mimesis (1950a, p. 220).[7] He agreed that heterosis explains the high frequency of alleles for sickle-cell anemia in Africa. However,

---

[7] Muller's rationale was that short-sightedness would have been maintained in primitive societies for its advantage in the performance of fine work by some of their members.

believing these to be exceptional cases, he contended that Dobzhansky's experiments in *Drosophila* had failed to control adequately for ordinary heterosis, that is, the concealing of deleterious recessives in the heterozygote. Muller understood overdominance to be a stop-gap evolutionary measure in the face of recent environmental change or unusual local conditions and believed that, over the long term, alleles that confer superior fitness only in heterozygotes would be replaced by those that exert similar effects in homozygotes: "this price [heterosis] is sometimes worth paying, when it gives us quickly what is much needed, and thus helps to tide the stock over until the gene in question can be "buffered," or until a more reliable one can be substituted" (1950b, p. 168). Dobzhansky conceded that, all things being equal, it is indeed advantageous for populations to contain alleles that confer maximal benefits to homozygotes, not heterozygotes. However, since all things are seldom equal, and, in particular, the constancy of environments cannot be assumed, it is advantageous for populations to be genetically heterogeneous and able to adapt to a wide range of environments.

3.1.2 Classical and Balance Conceptions of Genetic Mutation

Muller and Dobzhansky agreed that the vast majority of mutations — more than 99 percent according to Muller (1973c, p. 76) — that arise will be harmful to their bearers and eliminated through "negative" or "purifying" selection. Neither Muller nor Dobzhansky were bothered by what Muller referred to as "the seeming contradiction" that mutations furnish the "building blocks" of evolution and yet are harmful in the overwhelming majority of cases (ibid.). It is to be expected that most mutations would be harmful because organisms, as products of millions of generations of natural selection, are usually optimally adapted to their environments. "Random changes in any complex mechanism, such as a watch or an automobile, are more likely to injure than to improve it" (Dobzhansky 1955b, p. 107). Muller and Dobzhansky also agreed that beneficial mutations are exceedingly rare: "Consistently useful mutants are like needles in a haystack of harmful ones," wrote Dobzhansky (1962, p. 139). Muller estimated the occurrence rate of such mutations to be less than one in 10,000; he considered the one in 100 figure he used to calculate genetic load to be "very conservative" (1973b, p. 211). Although adaptive evolution depends on the occurrence of these exceedingly rare

beneficial mutations, this process is facilitated by changes in the environment because, once organisms are no longer optimally adapted for their environment, there is increased chance that a new mutation might prove beneficial. Over millennia, just as "watches and automobiles are changed for the better, step by step" (Dobzhansky 1955b, p. 107) with time, so too do organisms evolve. Dobzhansky, at least after the early 1940s, was sceptical that mutations could be neutral. Any change to the genetic structure was likely to have phenotypic effects and any phenotypic variation was likely to have effects on fitness. Muller, even arch-adaptationist that he was, allowed that mutations "of virtual indifference for survival" might arise, albeit at a very low frequency. Despite their low rate of occurrence, these could account for much of the "superficial genetic polymorphism" in human populations because, invulnerable to selection pressure, they accumulate over many generations (1950b, p. 142).[8]

Although Dobzhansky and Muller agreed that mutations are harmful in the vast majority of cases and only very rarely beneficial, Dobzhansky believed that Muller was wrong to consider mutation to be an "evil" that is necessary for the possibility of future evolution in a species. Due to the ubiquity of heterotic selection, Dobzhansky argued that mutation promotes the present fitness of a population:

> According to the balance hypothesis, the role played by mutation in the life of Mendelian populations appears in a new light. In order to preserve a high degree of fitness a population must contain a variety of alleles of many genes. This is true not only because mutation supplies the raw materials from which evolutionary changes can be compounded by natural selection. The role of mutation is important for present as well as future fitness. It is needed to maintain the species as it is today. If the fitness of a species depends to any appreciable extent on the presence of heterotic gene alleles.., there must be a source of supply of new alleles to replace those that become lost by chance or otherwise from the gene pool. Mutation is, then, not only the price for evolutionary plasticity; it is also the tax levied in order to preserve the status quo. (Wallace and Dobzhansky 1959, p. 165)

Associated with this specific disagreement over the relative importance of heterosis, Muller and Dobzhansky differed in their general willingness to label individual alleles as

---

[8] This mention of neutral mutation and "superficial genetic polymorphism" is very rare in Muller's writings which overwhelmingly emphasize the deleterious effects of even slight departures from the norm.

harmful, neutral, or beneficial across the board, that is, regardless of context. Muller stressed additive gene effects and the ubiquity of partial dominance (incomplete recessiveness): "Most mutant genes have a certain degree of dominance, usually enough to be 'effective'" (1950b, p. 173).[9] This makes it (approximately) twice as good or bad to have two copies of a gene than to have just one. For example, where two alleles are segregating at a locus, heterozygote fitness falls midway between the fitnesses of the two homozygotes. Although a particular genetic or environmental background may induce a strengthening or weakening of a gene's effects, its polarity remains constant. A "good" allele is always "good" and a "bad" allele is always "bad."

In contrast, with his emphasis on the prevalence of heterotic loci, as well as the significance of gene-gene and gene-environment interactions, Dobzhansky emphasized context. He considered gene effects to be nonadditive and the adaptive values of individual alleles always to be context-dependent. Dobzhansky's phrase "consistently useful mutations" is key to the source of his disagreement with Muller concerning the harmfulness of mutations. The "needle in the haystack" mutation that is advantageous in all genotypic and environmental backgrounds will be positively selected and become fixed in the population. Less infrequently, Dobzhansky believed, a mutation will be advantageous given some genotypic and environmental backgrounds and disadvantageous given others. Component parts of a system have properties only in virtue of the positions they occupy in the system. It is the "total constellation of genes" — at the same locus, in the rest of the genome, and across the population — that determines the adaptive value of any single allele. An allele never has value in itself: "every gene is potentially heterotic and potentially deleterious" (Wallace and Dobzhansky 1959, p. 164):

> we will not be justified in assuming (without sufficient evidence) that a mutant which is harmful or heterotic in certain combinations of genes will behave similarly in all genetic constitutions it encounters; some genes are known to interact favorably with some but unfavorably with other genes. (ibid., p. 162)

---

[9] To argue for the prevalence of partial dominance, Muller appealed to the phenomenon of dosage compensation. In the presence of complete dominance, there would have been no stimulus for the evolution of a system that compensates for the different dosages of sex-linked genes received by males and females: one dose would have been good enough for males (1950b, p. 129). Less detrimental mutants are even more likely than lethals to be partially dominant because they confer less impetus for complete dominance in the normal allele to evolve.

In a population that is genetically very heterogeneous, the "combining ability" or "co-adaptability" of an allele begins to matter more: "A genetic good mixer becomes superior to a genetic rugged individualist" (Dobzhansky 1955a, p. 3).

3.1.3 Classical and Balance Conceptions of Genetic Variation

Mutation is the ultimate source of genetic variation in a population, although, in sexually reproducing diploid species, new combinations of alleles arise through recombination. Not surprisingly, Muller and Dobzhansky viewed the adaptive significance of genetic variation in a population or species differently with respect both to the present and to the future. Muller's "classical" hypothesis on the genetic structure of populations predicts that, in a population under selection pressure, diploid genomes will be homozygous for the "normal" or "wild type" allele at almost all loci and the well-adapted population will therefore be genetically homogeneous. With ubiquitous partial dominance, natural selection is not only precise, but effective and rapid. At equilibrium, dominant deleterious alleles are eliminated from the population at twice the rate as recessives, yielding significantly less variation due to recurrent mutation, or mutational load, than would be the case if most deleterious mutants were completely recessive. We have see that, for Muller, even small degrees of variance in continuous traits have an impact on fitness. Muller believed the disadvantage conferred to be "roughly proportional" to the extent to which the trait deviates in an individual from the population mean: "there is no actual threshold amount of difference which suddenly emerges as disadvantageous" (1950a, p. 198). Hence, virtually all genetic variation in a population represents a "load" for the species. Fortunately, since at genetic equilibrium deleterious mutations are eliminated by natural selection at the same rate as that at which they occur, the "genetic load" placed on the species or population by recurrent mutation is maintained at a constant, tolerable level.[10]

Dobzhansky's "balance" hypothesis on the genetic structure of populations predicts that, in a population under selection pressure, the typical diploid genome will be heterozygous at most of its loci and a well-adapted population will be genetically

[10] The genetic load principle originated with J. B. S. Haldane in a 1937 paper.

heterogeneous. Dobzhansky held that the superior fitness of individual heterozygotes reflects their abilities to adjust to a wider range of environments than homozygotes. He believed that his *Drosophila* research had demonstrated that heterozygotes exhibit better homeostatic regulation of development under variable environmental conditions than flies homozygous for "wild type" alleles. Since no single genotype, heterozygous or homozygous, is advantageous in all environments that a population encounters, fit populations will include a variety of genotypes (1955a, p. 10). In his emphasis on the adaptive benefits of heterozygosity for individuals and populations alike, Dobzhansky was influenced by I. Michael Lerner's (1954) *Genetic Homeostasis*.[11] Lerner considered homeostasis to be a property of both individual organisms (developmental homeostasis) and Mendelian populations (genetic homeostasis). Lerner defined genetic homeostasis as "the property of the population to equilibrate its genetic composition and to resist sudden changes" (p. 2). He believed heterozygosity to be responsible for the homeostatic properties exhibited by both individuals and populations. Heterozygosity offers individuals "superior buffering capabilities" (p. 6). The "buffering" of individuals results in the "buffering" of populations: "This property of populations emerges from stabilizing [balancing] selection operating on individuals" (pp. 118-119).

Lerner argued that heterozygosity fosters the "successful existence" of Mendelian populations in two ways (p. 118). First, in the short term, under usual environmental conditions, heterozygosity ensures the stability of populations because "it permits a large proportion of individuals to exhibit combinations of phenotypic properties near the optimum" (p. 108). Second, in the long term, should environmental conditions change, heterozygosity provides populations with plasticity because the genetic variability that underlies phenotypic uniformity functions as "genetic reserves." Dobzhansky agreed with Lerner on both counts. Since mutations occur randomly with respect to their adaptive value, Dobzhansky considered it advantageous "for the species to possess at all times a

---

[11] Lewontin (1987) argues that Dobzhansky was influenced by Lerner in changing his position regarding heterosis from a view that considered it to be a relatively uncommon phenomenon associated with the chromosomal inversions that he had observed in natural populations of *Drosophila pseudoobscura* to a view that embraced it as a phenomenon ubiquitous in nature (p. 345).

store of concealed, potential variability" in order to accommodate future environmental changes (in Beatty 1987, p. 282). An allele neutral or harmful under present circumstances may well prove adaptive should circumstances change:

> This store will presumably contain variants which under no conditions will be useful, other variants which might be useful under a set of circumstances which may never be realized in practice, and still other variants which were neutral or harmful at the time when they were produced but which will prove useful later on. (ibid.)

Muller's opinion seems to have been that, despite the randomness with which mutations occur, the natural mutation rate is adequately high to sustain future evolution and that maintaining stores of deleterious mutations as "the price for evolutionary plasticity" levies too high a cost to the fitnesses of present populations. Muller admitted that, in the event of sudden environmental change, "genetic reserves" might "act as a damper to prevent a merely temporary selection from altering the population too hastily and so doing long-term damage greater than the short-term good" (1973a, p. 194). This is consistent with his belief in absolute fitness values for individual alleles: an allele once deleterious could never prove advantageous except in the short term.

These disagreements between Muller and Dobzhansky over the adaptive value of intraspecific genetic variation, the harmfulness of mutations, and the relative importance of partial dominance versus overdominance and "purifying" versus "balancing" selection are manifested in the "genetic load" controversy. Muller was extremely concerned about the threats to the "health" of the gene pool posed by the mutagenic effects of exposure to radiation as well as the long standing effects of civilization — improvements in medical care and sanitation, for example. At genetic equilibrium, deleterious mutations are eliminated from the population at the same rate as that at which they occur and remain, therefore, at a constant level. When natural selection is relaxed, or the mutation rate increases, equilibrium is upset and an excess of deleterious alleles accumulates in the gene pool. Muller estimated that "the average individual is probably heterozygous for at least 8 genes, and possibly for scores, each of which produces a significant but usually slight detrimental effect on him" (1950b, p. 170) "adding up to at least a 20% natural disadvantage" (ibid., p. 144). The effects of these partially dominant mutations are withstood only because, having evolved in primitive conditions, the germ plasm is on the

whole quite hardy and where it is deficient it can be propped up by modern technological and medical advances.[12] Muller stressed that one day the debt must be repaid and equilibrium restored. Maintaining faith that medical and technological progress will keep sufficient pace to accommodate an ever-increasing accumulation of mutations is like believing in the possibility of "push[ing] back the flowing waters of a river with one's bare hands" (ibid., p. 146). He painted an extremely grim picture of what lies ahead if society does not take action by limiting exposure to radiation and replacing natural with artificial selection. Early on, "people's time and energy ... would be devoted chiefly to the effort to live carefully, to spare and to prop up their own feeblenesses, to soothe their inner disharmonies and, in general, to doctor themselves as effectively as possible" (ibid.). Eventually, because natural selection is responsible for the maintenance of traits and not just their evolution — "it is, in a sense, only selection that holds the body in shape" (1973b, p. 227) — like animals living in caves who have lost the ability to see, our "natural biological organization" would yield to "mutation pressure" and disintegrate to be replaced in our descendants by "complete disorder" (1950b, p. 146).

Dobzhansky was not in entire disagreement with Muller's analysis. He believed the load principle itself to be correct. He rejected Muller's term 'genetic death' for the elimination of individual mutant alleles at a rate equal to that of their occurrence in populations at genetic equilibrium, but not the concept it expresses, affixing the "less dramatic" label 'genetic elimination' (1962, p. 290). Dobzhansky also recognized the importance of stabilizing (normalizing) selection in the maintenance of species form: "Normalizing selection opposes the spread in the populations [sic] of detrimental mutants.... It is obviously important in humans and other populations, since it prevents them from becoming arrays of freaks" (ibid., pp. 155-156). Dobzhansky took specific

---

[12] The complete passage reads:

> [It is] so fortunate for all of us in this generation, that our germ plasm was selected, in our more primitively-living ancestors, for a world without central heating or refrigerators, without labor-saving devices in the home, in industry or in agriculture, without sewers or bathrooms, and without knowledge of contraceptives, asepsis, antibiotics, calories, vitamins, hormones, surgery or psychosomatic treatment. And so now for the first time, with the newly found aid of all these devices and methods, the average American, in spite of his eight or more inborn disabilities, adding up to at least a 20% natural disadvantage, manages to get by for almost the ... "normal" [life] span. (Muller 1950b, p. 144)

issue with Muller over the calculations and dire prognostications that follow directly from Muller's assumptions that partial dominance and stabilizing selection prevail in nature. With his own money placed on overdominance and balancing selection, Dobzhansky believed that further knowledge was essential before policy initiatives concerning radiation exposure and eugenics were implemented. It was not opposition to policies in these areas *per se* that accounted for Dobzhansky's reluctance. Like Muller, he held that unnecessary increases in radiation exposure should be avoided and sought human control over evolution. For effective policies, however, better understandings of the genetic structure of populations and the significance of "genetic load" for species were necessary. For example, if the balance hypothesis is true, "instead of making everybody alike, possessing some one optimal genotype, [eugenics] will have to engineer a gene pool of the human population that would maximize the frequency of the fit and minimize that of the unfit" (ibid., p. 127). Dobzhansky also believed Muller to be mistaken in contrasting "genetic loads" in human and natural populations. According to Dobzhansky, natural and human populations bear similar "loads": "Man cannot blame his genetic load on his civilization, although civilization may well change its composition" (Wallace and Dobzhansky 1959, p. 159). Evolution in humans has not ceased; rather, it has become a product of both natural and cultural forces. For example, allelic variants that confer susceptibility to certain drugs are subject to increased selection pressure under conditions of civilization. Additionally, since the late 1920s, evidence had been convincing that natural populations of *Drosophila*, uniformly "wild-type" in appearance, actually conceal a great deal of genetic variation.[13] Flies that appear "wild-type" are "rarely, if ever" "free of deleterious genes of all sorts." This should, on Muller's view, represent an "enormous" genetic load and yet the species is "flourishing" (ibid., p. 117).

3.1.4 Classical and Balance Conceptions of Genetic Normality

Muller believed that stabilizing selection maintains a single optimal or "normal" allele at

---

[13] This research was carried out by Chetverikov, Timoféeff-Ressovsky, and Dubinin and his collaborators from the mid-1920s to the early 1930s. The method was to capture *Drosophila* in the wild and then cross them in the laboratory to reveal the presence of hidden recessive mutations (Dobzhansky 1955a, p. 4).

almost all loci and preserves a "normal" species-type: "in the great majority of cases it is after all valid to speak of a 'normal gene' and a 'normal type.' This gene or type can vary only within very narrow limits of effect without a significant reduction in the average over-all fitness of the organism" (Muller in Crow 1987, p. 377). Since there is a single optimal or "normal" allele at each locus, a single optimal or "normal" genotype underlies the "type" specimen. As complete dominance is rare and overdominance is aberrant and temporary as well as rare, the ideal diploid genotype is composed of two identical haplotypes with the optimal or "normal" allele present at each locus. Given a single optimal genotype for a species or population, normally distributed values for continuously varying organismal traits come to be understood in a specific way. Stabilizing or normalizing selection maintains continuously varying traits at their optimal values by eliminating alleles associated with increased variance of a trait due to the non-directional forces of evolution — random mutation, migration, and drift. Since even slight departures from the norm are maladaptive, says Muller,

> we become aware of the falsity of the assumption so often made, by both biologists and medical men, which holds that variants within the so-called "normal range" (i.e. those falling within, say, the middle 80 or 90 per cent of the area of the curve of variation) are in effect "normals," possessing no or negligible disadvantage. (1950a, p. 218).

Under selection pressure, the mean value for a trait represents its optimal value and deviation from the mean indicates the presence of inferior alleles at relevant loci.

Muller essentially understood the normal distribution of values for continuously varying traits to represent the cumulative expression of the interplay of two types of alleles at each locus: a single optimal allele and a number of inferior mutant variants. He did not conceive the normal distribution of values for a trait in terms of the interaction of numerous alternate but acceptable alleles at a collection of loci, all with varying degrees and directions of effect on the trait's expression. The familiar bell-shaped normal distribution curve for continuously varying traits is maintained because mutations of small effect both occur at higher rates than those of large effect and, since they are eliminated more slowly by natural selection, accumulate to higher frequencies in the population. However, because stabilizing selection maintains the frequencies of mutant alleles at significantly lower levels than normal alleles, "the average grade of the character, the

norm, [is] rather well defined and comparatively stable" (ibid.). Should selection pressure decrease and mutation pressure increase — due to the effects of civilization and exposure to radiation, for example — the distribution curve will flatten and variance will increase, leaving the norm less well defined.

For Dobzhansky, unlike Muller, variation is not deviation from an adaptive norm; rather, variation *is* the adaptive norm. Just as Dobzhansky was less willing than Muller to refer to alleles as beneficial, neutral, or harmful regardless of context, he was also less inclined to designate normal alleles or to delineate a normal species-type. Dobzhansky claimed that no single genotype can be considered to be normal for a species. Studies of natural populations of *Drosophila* beginning with those of Sergei Chetverikov and his fellow Russian researchers in the 1920s had revealed the presence of a great deal of concealed genetic variation in flies uniformly wild-type in phenotypic appearance. Dobzhansky extended these studies in his own work on chromosomal inversions in natural populations of *Drosophila pseudoobscura* and was led to conclude that the concept of "wild-type," taken to refer to genotype not phenotype, is invalid: "Wholly homozygous and mutant-free men, or cats, or mice, or Drosophilae, have never existed in nature" (Wallace and Dobzhansky 1959, p. 159).

Instead, Dobzhansky adopted the concept of a species adaptive norm that represents "a great array of genotypes, not just one or a few genetic complexes" (1962, p. 127). This "array" consists of "related genotypes consonant with the demands of the environment" (1955a, p. 3). The genotypes worthy of inclusion in the species adaptive norm can be identified on the basis of their "norms of reaction." "Norms of reaction" express the range of phenotypes that a given genotype exhibits in different environments. Dobzhansky granted that "the boundary between the adaptive norm and the genetically handicapped sector of the population is not sharp" (1962, p. 127):

> One possible definition of the adaptive norm might exclude only those persons who, because of their genetic defects, must be permanently hospitalized or cared for in special institutions; another definition would exclude even those whose genetic handicaps require attention or special regimens at any time in their lives. (ibid.)

Dobzhansky emphasized that normality and abnormality are statistical terms that are properly used only in reference to a genotype's frequency in the population. As a result

of natural selection, the relative fitnesses and frequencies of individual genotypes do tend to coincide, although this is true only in the usual environmental conditions. The "norm of reaction" provides a better indication of the adaptive value of a given genotype than does its frequency because it is a functional measure that ranges across different environments. A similar relationship between frequency and fitness does not hold for individual alleles, however, because there is no consistent correlation between their fitnesses when homozygous and when heterozygous (1955a, p. 5).

Muller and Dobzhansky differed also in their views on the relationship between individual and population genetic norms. Muller rejected Lerner's account of genetic homeostasis, considering it to be "an essentially mystical doctrine, representing a revival from pre-Mendelian times" when the particulate nature of the gene was yet unknown (1973b, pp. 225-226). Muller argued that properties that attach to gene loci, genotypes, and Mendelian populations merely reflect the additive effects of their component parts — the individual alleles — and not how these parts are arranged. For Muller, terms like "homeostasis," "adapted," and "adaptable" properly describe individual organisms, not entire populations. For Dobzhansky, on the other hand, it makes sense to talk about populations, and not just organisms, as adapted or adaptable. These properties may not coincide. At heterotic loci, for example, the fitness of individual homozygotes is sacrificed for the fitness of the population as a whole. Dobzhansky, like Lerner, emphasized the emergent properties of populations. The Mendelian population represents "a level of organic integration," Dobzhansky wrote, "which obeys its own laws and contains its own regularities" (1955a, p. 14). Populations have properties that emerge from the arrangements of their parts:

> A gene system may be likened to a mosaic picture, and the genes to the component stones. The nature and quality of a mosaic picture are determined obviously by the pattern in which the stones are placed, as well as by the characters of the separate stones. (Dobzhansky 1955b, pp. 175-176)

Evolutionary change involves "a re-patterning of the gene pool" (Dobzhansky 1955a, pp. 3-4). This "corporate genotype" (ibid., p. 12) comprises individual genotypes, not individual alleles. This is because of the prevalence of nonadditive interactions between alleles at the same, and at different, loci. In Lerner's words: the "totality of interaction

between all components of a genotype forms a more important selection criterion in nature than the additive properties of single genes" (1954, p. 119).

3.2 The Neutralist-Selectionist Debate and "Non-Darwinian" Evolution

When Motoo Kimura introduced the neutral theory of molecular evolution in 1968, he did so in opposition to what he perceived to be the prevailing neo-Darwinian panselectionist "consensus" reached by the 1960s that "every biological character can be interpreted in the light of adaptive evolution by natural selection" and that "almost no mutant genes are selectively neutral" (1982, p. 4). One year later, J. L. King and T. H. Jukes published their own account of neutral evolution under the title "Non-Darwinian Evolution." These characterizations are misleading, however, because, in the *Origin of Species*, Darwin allowed for neutral variation in traits as well as the possibility that such variation could one day become adaptive. He also indicated that once-adaptive traits would become increasingly variable when selection ceases (Crow 1985, p. 3). During the 1920s-1930s, genetic drift, also called the "Sewall Wright effect," was believed responsible for many subspecies and species differences, especially among systematists and evolutionists in the United States. However, it is certainly the case that, during the 1940s and early 1950s, there was a "hardening of the evolutionary synthesis" (Provine 1986, p. 404; Gould 1983). Empirical evidence became available that traits regarded only a few years earlier as selectively neutral (for example, chromosomal inversions in *Drosophila* and the human blood groups) were adaptive. It is at this molecular level that Kimura's challenge rests. The neutralist-selectionist debate concerns the selective value of *molecular*, and not gross phenotypic, characteristics. It is generally agreed that substantial morphological variation is correlated with fitness. I begin this section by outlining the terms of the debate. I then compare neutralist and selectionist outlooks on genetic mutation, genetic variation, and genetic normality.

3.2.1 The Terms of the Debate

Like the classical-balance debate, the neutralist-selectionist debate concerns the relative

importance of different evolutionary forces, and not whether they occur at all. The question at issue in the neutralist-selectionist debate is whether natural selection or genetic drift predominates in evolutionary changes at the molecular level. 'Molecular' here refers to proteins and DNA and the variability in their respective amino acid and nucleotide sequences. At the level of gross phenotypic changes in form or function, Kimura was as ardent a selectionist as any. Like the neo-Darwinists, he accepted that phenotypic evolution is adaptive and proceeds through the positive selection and gradual incorporation of very rarely occurring beneficial mutations. What the neutral theory of molecular evolution did was to drive a wedge between molecular and phenotypic evolution. These became viewed as distinct processes governed by different "laws." Deterministic changes at the level of the organism result from natural selection. Stochastic changes at the molecular level are due to genetic drift, their rate determined by "the structure and function of molecules and not by environmental conditions" (Kimura 1979b, p. 104):

> The laws governing molecular evolution are clearly different from those governing phenotypic evolution. Even if Darwin's principle of natural selection prevails in determining evolution at the phenotypic level, down at the level of the internal structure of the genetic material a great deal of evolutionary change is propelled by random drift. Although this random process is slow and insignificant in the time frame of man's ephemeral existence, over geologic time it makes for change on an enormous scale. (ibid., p. 106)

Only at the molecular level, therefore, was Kimura extending Wright's theory of drift from small to large populations.

Kimura identified two parts to the neutral theory. The first part concerns the substitution of nucleotides (or amino acids) in evolution. Kimura contended that "a majority of nucleotide substitutions in the course of evolution must be the result of random fixation of selectively neutral or nearly neutral mutants rather than positive Darwinian selection" (1982, p. 7). For support, he appealed to two main observations: the approximately constant rate of evolution in terms of amino acid substitutions per year in different lineages — in the hemogoblin molecules of humans and carp, for example — and the higher rate of evolution in molecules or parts of molecules subject to less functional constraint (1979b, pp. 102ff). The second part of the theory concerns molecular variation within as opposed to between species. Kimura believed intraspecific

DNA and protein polymorphisms to reflect the transient manifestation of the random drifting of neutral alleles on their way either to eventual fixation or loss from the population: "many of the enzyme polymorphisms are selectively neutral and maintained by the balance between mutational input and random extinction" (1982, p. 7). Hence, this second aspect of the neutral hypothesis, intraspecific polymorphism, is embedded in the first, the molecular differentiation of species.

Kimura's neutral theory was motivated, in part, by the 1966 discoveries of high levels of protein polymorphism in natural populations.[14] These studies using gel electrophoresis were carried out in *Drosophila* by J. L. Hubby and R. C. Lewontin (1966) and Lewontin and Hubby (1966) and in humans by Harry Harris (1966). It had been believed that such determinations would resolve the classical-balance debate which centred, as we have seen, on the amount of genetic variation that is present in natural populations. However, when field data were finally obtained, the findings did not entirely resolve the classical-balance debate but instead presented additional new problems. The observed levels of protein polymorphism far exceeded the predictions of the classical school.[15] This meant that partial dominance could not be ubiquitous. The observed variability could not reflect the retention of deleterious mutant alleles because this "mutational load" would place too much strain on a population's fitness. Were the balance position true, given the amount of genetic variation, populations would still face an intolerably high "load" — in this case, a "segregational load" due to the inferior fitness of homozygotes. Nor could the polymorphisms be explained in terms of the gradual incorporation of new advantageous alleles by directional selection since the "substitutional load" ("cost of selection") associated with the elimination of their predecessors would also be too great (Kimura 1968). Kimura presented a solution that was unanticipated by players on either side of the classical-balance debate. If the protein polymorphisms

---

[14] See Dietrich (1994) for a broader historical overview of the origins of the neutral theory.

[15] Crow (1987) states that, in 1966, although the amount of heterozygosity found in natural populations fell somewhere between the values expected by both the classical and balance camps, the data were, on the whole, more consistent with Dobzhansky's position. Crow argues, however, that improved methods since 1966 have brought estimates of heterozygosity in natural populations closer to the levels Muller's hypothesis predicts.

observed in natural populations are selectively neutral, their frequencies governed neither by "purifying" nor balancing selection but by drift and mutation alone, there is no "load" of any type to be borne by the population.[16]

Selectionists contend that variability in protein and DNA structure is primarily adaptive, both between and within species. Nucleotides and amino acids are substituted in the evolution of a species because they are selectively advantageous. Selectionists are sceptical that variation in proteins, in untranslated regulatory DNA and introns, in translated but "silent" nucleotide substitutions, and even in the large proportion of the genome often referred to as "junk" DNA that has no known coding or regulatory function, makes no difference to fitness. On this view, intraspecific polymorphisms are not transient manifestations of passively drifting neutral alleles but are actively maintained by balancing selection. Several different balancing mechanisms are possible: heterozygote advantage, frequency-dependent selection, cyclical fluctuations in selection coefficients due to environmental conditions that vary over time (especially over the life-cycle), geographically diverse environmental conditions, habitat selection, etc. Selectionists have challenged Kimura's interpretation that the high levels of protein polymorphisms observed in natural populations can only be explained by their selective neutrality because of the high mutational, segregational, and substitutional loads that they would otherwise present. One way has been to explain the observed levels of protein polymorphisms in terms of nonheterotic forms of balancing selection that do not present similar problems of segregational load — for example, frequency-dependent selection. Another way has been to appeal to truncating selection. Several authors (King 1967; Milkman 1967; Sved, Reed, and Bodmer 1967) responded independently to the presentation of the problem of segregational load in Lewontin and Hubby (1966) to point

---

[16] Lewontin (1987) considers the failure to resolve the classical-balance debate on the basis of the evidence of extensive genetic variation in natural populations and its subsequent transformation into the neutralist-selectionist debate to be the result of a conflation of two questions as the result of a missed premise. One question concerned monomorphism and polymorphism and whether the production of new mutations or existing genetic variation serves as the rate limiting factor in evolution. The other question addressed the classical and balance positions on the relative importance of "purifying" and balancing selection. The missed premise is that selection operates on the genetic variation present in a population. Kimura challenged this assumption shared by both Muller and Dobzhansky.

out that Lewontin and Hubby's calculation that the maximally fit completely heterozygotic female *Drosophila* in a population that is maintaining its present size would have to lay 10^{43} eggs assumes multiplicative selection where an individual's total fitness is the product of the fitness coefficients at each loci considered separately. If, as in truncating selection, a threshold for selection is assumed instead, segregational load poses less of a problem.

Although Kimura's neutral theory is about molecular evolution through random drift, his appeal to functional constraints on genome evolution forces attention to natural selection's effects at the molecular level and how these relate to phenotypic changes in evolution. Kimura distinguishes between positive and negative selection. Positive directional selection is responsible for the evolution of phenotypic form and function; however, because the incorporation of a newly favourable mutant allele is such a rare event, a theory of molecular evolution can easily afford to ignore positive selection's effects on the genome. The neutral theory assumes that, where natural selection operates, it is a stabilizing force that preserves phenotypic form and function through the elimination of deleterious mutations: "It is known, since the great work of Muller in the early days of *Drosophila* genetics, that negative selection is the most common form of natural selection," writes Kimura (1982, p. 12). In functionally important areas of the genome, mutations are likely to be deleterious and eliminated by negative selection. As a result, these regions will be highly conserved in evolution and vary little either within or between species. In functionally unconstrained regions of the genome, selection pressure is eased and mutations are likely to be neutral or nearly neutral in their effects. Molecular evolution is overwhelmingly due to the chance fixation of nucleotides in such regions.

3.2.2 Neutralist Conceptions of Genetic Mutation

Kimura diverges from both Muller and Dobzhansky in emphasizing the prevalence of selectively neutral alleles. It is important to clarify what Kimura meant by selective neutrality. Although it is sufficient for neutrality that alleles not contribute to function — as may be the case for "junk" DNA, for instance — it is not necessary: "The neutral theory ... does not assume that neutral genes are functionless but only that various alleles

may be equally effective in promoting the survival and reproduction of the individual" (Kimura 1979b, p. 100). Neutrality is therefore a comparative notion. Physiologically, at the level of the individual organism, a neutral allele is a variant that is indistinguishable from the "wild-type" allele in terms of its phenotypic effects. In other words, alleles with functionally equivalent effects are neutral. A nucleotide substitution that is "silent" and results in no amino acid substitution in the encoded protein because of the redundancy of the genetic code is likely to be neutral. A nucleotide substitution that does result in the substitution of an amino acid, but one that is similarly charged or in a functionally unimportant area of the protein, may also be neutral. However, although functional equivalence is sufficient for neutrality, it is not necessary. For the population geneticist, at least if she or he is a neutralist, "this equality need not be perfect" (Kimura 1982, p. 11). Neutral alleles are those whose differential effects on fitness are sufficiently small that their behaviour in a population depends on chance and not on natural selection:

> The essential part of the neutral theory is not that the alleles involved are selectively neutral in the strict sense. Rather, the emphasis is on mutation and random drift as explanatory factors in molecular evolution because the selection intensity involved is exceedingly small. (ibid., p. 49)

Neutrality depends on population size as well as on fitness coefficients. An allele that is neutral in a small population may be subject to selection, positive or negative, in a large population.[17]

When Kimura introduced the neutral theory in 1968, he emphasized the high rate of occurrence of neutral and near-neutral mutations. He suggested that such mutations represent the greatest proportion of mutations that arise: "the very high rate of nucleotide substitution which I have calculated," Kimura wrote, "can only be reconciled with the limit set by the substitutional load by assuming that most mutations produced by nucleotide replacement are almost neutral in natural selection" (1968, p. 625). In 1979, Kimura modified his original theory. The new "effectively neutral mutation model"

---

[17] Estimates of the relationship between fitness and population size range from a neutral allele having a difference in fitness "much smaller, such as, less than 10% of the reciprocal of the effective population size" (Kimura 1982, p. 11) to "less than the reciprocal of four times the effective population number" (Crow 1981, p. 5) to "smaller than the reciprocal of the effective population number" (Crow 1972, p. 307).

emphasized deleterious over neutral mutations and assumed "that molecular evolution and polymorphism are caused by random drift of very slightly deleterious but effectively neutral mutations" (1979a, p. 3444).[18] This revision was inspired by Tomoko Ohta's 1973 model in which she proposed that the majority of "neutral" alleles are not strictly neutral but very slightly deleterious.[19] Whereas Ohta's model entails that all mutations in very large populations would be in mutation-selection balance, Kimura's model accommodates mutations that are neutral regardless of population size. This leaves room for neutral molecular evolution to occur in large and small populations alike.[20] Kimura estimated that neutral mutations occur at 14 percent of the total mutation rate, although the exact rate would depend on the degree of functional constraint operating at a locus (1983a). That the majority of mutations are harmful is to be expected because organisms are already well-adapted to their environments. Unless only mildly deleterious, these mutations are eliminated by negative selection and do not accumulate in the population. The neutral theory of molecular evolution excludes entirely the class of beneficial mutants: "Advantageous mutations may occur, but the neutral theory assumes that they are so rare that they may be neglected in our consideration" (Kimura 1991, p. 5).

As in the classical-balance controversy, dissension among neutralists and selectionists involves neutralists' willingness to label alleles good, bad, or indifferent in a way that seemingly disregards context. Bruce Wallace (1991) argues that neutralists have abandoned a strict, though admittedly "overly stringent" and "unverifiable," notion of neutrality as physiological indistinguishability for a pragmatic one: "in their view, a neutral allele is one that behaves *as if* it were neutral" (pp. 146, 152; my italics):

---

[18] Takahata (1994) notes that, in post-1986 publications, Kimura no longer mentions the "effectively neutral model" but returns to the original neutral theory and its dichotomy between "completely neutral" and "definitely deleterious" mutations (p. 562).

[19] These "slightly deleterious" and "effectively neutral" models permitted resolution of a problem for the neutral hypothesis: that observed levels of heterozygosity in large populations, at 0-20%, were significantly lower than predicted.

[20] Kimura modified Ohta's model by assuming a gamma distribution of selection coefficients for mutants, rather than an exponential distribution. He defined effectively neutral mutants as those with selection coefficients less than 1/(2N) and selective neutrality as the limit in which "the selective disadvantage becomes indefinitely small" (1979a, p. 3440).

> [Alleles] are neutral if their average fitnesses are equal (in large populations) or nearly so (in small populations). They are neutral if their fitnesses fluctuate with frequent reversals either through time, within patchy environments, within different background genotypes, or any combination of these three factors. (p. 146)

However, once neutrality in the strict sense is abandoned, as Wallace notes, it cannot be taken for granted that selection is not operating: "Selection pressures that fluctuate through time or that vary depending upon the individual's situation result in apparent neutrality" (p. 152). But "randomness does not imply an absence of selection (i.e., neutrality in a strict sense); on the contrary, it may imply a multiplicity of selections that in toto generate effects that appear to be random" (p. 147).

Whereas selectionists like Wallace stress that the fitness coefficients of alleles fluctuate depending on their genetic and environmental backgrounds, neutralists recognize context-dependence only due to population size: a slightly deleterious mutant allele behaves *as if* it is neutral in a small population, but responds to selection in large populations. As we have seen, the neutralist discounts the importance of all forms of balancing selection. Hence, the selective value of a particular genotype does not depend on its relative frequency in the population as is the case for frequency-dependent selection. Nor does the selective value of a given allele depend on other alleles that are present at the same locus as is the case for heterotic selection. Neutralists also dismiss the significance of variable environments on the fitnesses of individual alleles and genotypes in a population by appealing, as did Muller, to the canalizing effects of stabilizing selection: "In higher organisms particularly, homeostasis counteracts external environmental changes just as it does internal physiological changes; fluctuations in the environment do not necessarily imply comparable fluctuations in the Darwinian fitness of mutant genes" (Kimura 1979b, pp. 100, 102). The neutralist treats selection coefficients for individual alleles as fixed even across geological time: estimates of species divergence times based on the neutral molecular clock assume that amino acid substitutions have occurred at a constant rate and that selection pressures and environmental conditions can be ignored.

3.2.3 Neutralist Conceptions of Genetic Variation

The neutralist understands the adaptive value of genetic variation to be a function of the

specific region of the genome in which it is found. Kimura's original neutral theory divides the genome into regions of two types: functionally constrained and functionally unconstrained. The small amount of variation found in functionally constrained regions reflects the presence of harmful mutant alleles in mutation-selection balance that are awaiting elimination by negative selection. The large amount of variation found in functionally unconstrained regions reflects selectively neutral mutant alleles, the frequencies of which depend on mutation rate and random drift. Genetic variation is therefore also of two types: harmful deviation ("genetic load") and selectively neutral difference (Mayr called this "evolutionary noise"). In Kimura's subsequent "effectively neutral" model, there is a less dichotomous treatment of variation. It is assumed that negative stabilizing selection operates continuously and to varying degrees across the genome and that the entire genome is therefore more or less functionally constrained. Consequently, strictly neutral mutations are assumed to be rare. The vast majority of mutations are deleterious, with chance prevailing over negative selection as degree of deleteriousness and population size decrease.

These shifts in the neutral theory from "strict" to "pragmatic" neutrality, neutral to slightly deleterious mutant alleles, and discontinuously to continuously acting selective forces on the genome result in changes in the neutralist's understanding of the adaptive value of genetic variation. It contributes to reconciling the neutral theory's divergent treatments of evolution at the phenotypic and molecular levels; as Kimura admits, the question "why natural selection is so prevalent at the phenotypic level and yet random fixation of selectively neutral or nearly neutral alleles prevails at the molecular level" cannot be ignored (1982, p. 48). Were all mutant alleles either strictly neutral or definitely deleterious, adaptive phenotypic evolution would be entirely dependent on the occurrence of new beneficial mutations, likely in novel environmental circumstances. With the change from "strict" to "pragmatic" neutrality, the continuously varying phenotypic traits that underlie much neo-Darwinian evolution find a genetic basis in nearly-neutral polymorphisms. Like Muller, Kimura emphasized the "stabilizing" or "normalizing" aspects of negative selection that preserve population mean values for continuously varying traits by eliminating deviate phenotypes. Unlike Muller, Kimura held substitutions and polymorphisms at quantitative trait loci to be the result of drift and

not positive selection given that, "if a large number of segregating loci or sites are involved in a quantitative character, the average selection coefficient per mutant under stabilizing selection may be exceedingly small" (ibid.).

Kimura's "effectively neutral" model understands the genetic variation that underlies continuously varying phenotypic traits as harmful deviation, as Muller did, rather than as innocuous difference. This is because the "new neutralism," as it is described by James F. Crow, is "a theory of substitution of mildly deleterious alleles, especially in smaller populations" (1981, p. 9). These "mildly deleterious" alleles are subject to stabilizing selection but, as population size decreases, chance effects predominate. That "effectively neutral, but, in fact, very slightly deleterious mutants accumulate continuously in every species" (Kimura 1979a, p. 3444) suggests that, over time, the genetic quality of a population gradually deteriorates. Kimura's estimate of the rate of loss of fitness per generation is 10^{-7} (ibid.). Although Kimura expressed optimism that this deterioration would in the long term be offset by the very rare "adaptive gene substitutions that must occur from time to time (say once every few hundred generations)" (ibid.), Crow (1972) points to its eugenic significance over the short term. Crow maintains Muller's focus on the importance to human welfare of continuously varying phenotypic traits and slightly deleterious alleles. Like Muller, he stresses that mutations of small effect arise more frequently and are found at higher levels in the population than mutations of large effect. Again, like Muller, Crow emphasizes the additivity of these "slightly deleterious" or "nearly neutral" alleles and their partial dominance in heterozygotes: "their effect on the population and on individuals in the population is one of mild weakening, which becomes important as the number of such genes increases" (p. 314). Society can ill afford to overlook the effects of the "slightly deleterious/nearly neutral" alleles that are responsible for variation in the quantitative (polygenic) traits that constitute the bulk of human variability.

Generally, for the neutralist, therefore, genetic variation is overwhelmingly bad (at least slightly), occasionally indifferent, and rarely, if ever, good. However, from the perspective of Kimura's "effectively neutral" model that assumes "pragmatic" rather than "strict" neutrality, existing genetic variation in a population is recognized to be of potential adaptive value in the present as well as in the future. Should environmental

conditions change, genetic variation may become redistributed in order to "track" the change:

> From time to time, the position of the optimum of a phenotypic character shifts due to change of environment, and the species tracks such a change rapidly by altering its mean. During this short period of change, extensive shift of gene frequencies is expected to occur at many loci. (Kimura 1982, p. 49)

Kimura did emphasize that directional selection of this type is infrequent and seldom causes gene substitutions: "most of the time, stabilizing selection predominates, under which 'neutral evolution' or random fixation of mutant alleles occurs extensively" (ibid.). He also allowed, however, that a population's "store" of neutral variants may prove useful for future adaptive phenotypic evolution. In novel environments, previously "neutral" or slightly deleterious alleles may prove beneficial and so alleviate the need to await for a rare beneficial mutant to arise:

> Sometimes, it is remarked that neutral alleles are by definition not relevant to adaptation, and therefore not biologically very important. I think that this is too short-sighted a view. Even if the so-called neutral alleles are selectively equivalent under a prevailing set of environmental conditions of a species, it is possible that some of them, when a new environmental condition is imposed, will become selected.... I ... believe that 'neutral mutations' can be the raw material for adaptive evolution. (1986, p. 345)

3.2.4 Neutralist Conceptions of Genetic Normality

Kimura's two neutralist models make different assumptions about genetic normality. His early model, which allows for a significant rate of occurrence of "strictly neutral" mutations, assumes that more than one allele at a given locus can be functionally "normal" or "optimal." But because it takes a very long time for a selectively neutral mutant allele to accumulate to an appreciable frequency in a population, polymorphic alleles at a locus are likely to be very old, perhaps older than the age of the species (Kimura 1983b, p. 215). In such cases, it is impossible to distinguish the mutant from the original allele at the locus (ibid., p. 217). Kimura's later "effectively neutral" model assumes that there is a single optimal allele at each locus. Insofar as "selective neutrality is the limit when the selective disadvantage becomes indefinitely small" (ibid., p. 241), strictly neutral alleles, absolutely equivalent to the original (optimal) allele at the locus, do not exist. In the "effectively neutral" model, deleterious mutant alleles are increasingly

likely to behave *as if* neutral as population size, degree of deleteriousness, and amount of functional constraint all decrease. Both versions of the theory assume that adaptive substitutions are so rare that they can be completely ignored. Hence, the "original" allele is assumed to be optimal. On the early version of the theory, it may have selectively equivalent successors; on the later version of the theory, all succeeding alleles are inferior.

To say, as Kimura does, that a mildly deleterious allele behaves *as if* it is neutral assumes that there is an arena other than the population in which the allele exhibits selective neutrality where its "true" selective value — its deleteriousness — can be identified. In other words, as John H. Gillespie (1991) points out, there exist absolute fitnesses from which relative fitnesses are derived (p. 265). The arena of absolute fitnesses is not actual; rather, it is the finite but very large population idealized in the theoretical population geneticist's mathemetical models. These absolute theoretical fitnesses must, however, have empirical content if we are to take seriously the possibility raised by Kimura that the continuous accumulation of "effectively neutral, but, in fact, very slightly deleterious mutants ... [may constitute] a threat to the survival and welfare of the species" (1979a, p. 3444). Many "optimal" (and "original") human alleles will be shared in common with related species because of functional constraints. Kimura notes that the molecules that the neutral theory is concerned with "must have had their essential designs perfected very far in the past" (1983b, p. 115). However, this leaves us facing a scenario eloquently captured by Crow: "It is as if evolution were steadily running down hill, as if gene functions were being successively inactivated. It is hard to think of oneself as an inactivated amoeba" (1981, p. 9). The neutral theory completely ignores the molecular changes that underlie the evolution of phenotypic form and function and deals only with neutral nucleotide and amino acid substitutions that have nothing to do with why we are humans and not amoebae.

Although adaptive evolutionary changes lie outside of the domain of Kimura's theory, the neutral theory, with its emphasis on stabilizing selection, does explain the maintenance of phenotypic form and function in species. Recall Kimura's argument that substitutions and polymorphisms at quantitative trait loci are the result of drift because, where a number of loci constribute to a trait, individual selection coefficients are low.

This is the basis for Roger Milkman's (1985) definition of the "mature species" as one where "the phenotype is at equilibrium, and the genotype is in a steady state" (p. 66):

> Phenotypically, the species remains constant, as intermediate values are favored. A balance is struck by the tendency of phenotypic variance to increase on one hand due to recombination and mutation, and to decrease on the other hand, due to selection. But the *genotype* is not stabilized. In the mature species, all the accepted alleles at a locus are liberated from selection. They become neutral. (pp. 67-68)

Milkman believes that the "mature species" can serve as a "reference standard" although, he admits, "we may not always know whether a given component of an optimal genotype is one of the very alleles favored by directional selection when the present phenotype evolved, or whether it is a recent substitute" (p. 80). Therefore, on Milkman's interpretation, the "optimal" genotype comprises alleles that were positively selected for early on in a species' evolutionary history or neutral substitutions that have since arisen. This does not accord fully, however, with Kimura's "effectively neutral" model where a mutant substitution will always be at least slightly inferior to the original allele at the locus. The "optimal" genotype, on this view, would comprise only alleles that were positively selected for early on in a species' evolutionary history or in an ancestral species' evolutionary history. Gillespie (1991) argues that Kimura can legitimately assume that all mutations will be slightly deleterious only if "selection has proceeded for a very long time — at least tens of millions of years — in a constant environment so that the most-fit allele has had an opportunity to displace the others" (p. 268). In any case, the "optimal" genotype is very old indeed.

3.3 The Human Genome Project and Its Evolutionary Context

In this final section of the chapter, I evaluate the evolutionary critique of the HGP as a throwback to pre-Darwinian, Platonic, essentialist, and typological thinking. I begin by summarizing the evolutionary beliefs that are implicit in the HGP's approach to genetic variation. I do so in reference to the differing conceptions of genetic mutation, genetic variation, and the "normal" genome implicated in the classical-balance and neutralist-

selectionist debates. I argue that the HGP shares a set of evolutionary beliefs that one would associate with the classical and neutralist sides of each of the respective debates and briefly discuss this continuity in terms of R. C. Lewontin's characterization of the neutral theory as "neoclassical." Finally, I present Dobzhansky's criticisms of Muller's classical approach to the genetic structure of natural populations as pre-Darwinian, Platonic, essentialist, and typological and address the question whether and why similar criticisms might apply to the HGP.

3.3.1 The Evolutionary Assumptions of the Human Genome Project

Recall, from Chapter Two, Walter Gilbert's (1992) characterization of molecular biologists' views on species. For Gilbert and his fellow molecular biologists, there is a "type organism" that exemplifies or defines the species which can itself be defined by its DNA. Is the DNA of the "type" organism an *example* of a sequence that is typical of the species and could realistically belong to some existing member of the species? Or is the DNA of the "type" organism *exemplary* of the species — that is, does it define a functional or ideal norm? If, as Gilbert contends, differences in genes define genus, family, and species groups, it seems that for cross-species comparisons the DNA sequence of any typical human would do. This definition of the human species by contrast to nonhuman primates involves a small proportion of total DNA — for example, it is estimated that humans and chimpanzees differ in only one out of every 60 nucleotide bases. If we are to accept with Gilbert that a "type" organism is "defined" by its DNA, this "definition" will have to include far more than simply nucleotides that are "unique" to the species. The DNA "definition" of an organism of one species "type" must include DNA that is shared by species of all types, ranging from "*E. coli* to the elephant." In this way, Gilbert's characterization of a sharp boundary between species is misleading: "Molecular biologists generally view the species as a single entity, sharply defined by a set of genes and a set of functions that makes up that entity" (1992, p. 84). Molecular biologists tend to be interested in *all* physiologically relevant genes and cellular processes and not just in some small "definitive" set of them. The reference sequence for the human genome is expected to represent a functionally normal or healthy genome that is

helpful for within-species, more so than cross-species, comparisons. Clinical researchers use reference sequences for specific genes to identify mutations that are associated with disease. As we saw in Chapter Two, in its use, the mtDNA reference sequence has received authoritative status as a representation of what is healthy or proper and not merely what is typical. As we will see in Chapter Four, there is less of a split in the aims of basic and clinical researchers than one might imagine. The discovery of normal gene functions is dependent on knowledge of their malfunctions, that is, their role in disease. Both basic and clinical researchers are interested in a reference sequence that represents a functionally normal genome — a prototype of human health, in other words.

Given that a *single* reference sequence is authoritative for the species as a whole, all human genetic variation is suspect. Any divergence from the reference sequence calls for assessment in order to determine its significance. Insofar as the reference sequence represents health and normal functioning, divergence falls into two possible, mutually exclusive, categories: the divergence constitutes *either* functionally irrelevant difference *or* deviation that is indicative of abnormal (impaired) function or disease. Since there is a *single* functional norm, were some divergence from the reference sequence be found to represent a deviation that improves upon function, the reference sequence is likely to be modified to reflect this. The question under consideration is: what sort of evolutionary assumptions are embedded in molecular biology's commitment to a single reference sequence that represents health and normal functioning for all humans?

Certainly, the notion of a single functional norm is consistent with Muller's classical theory of the genetic structure of populations. On Muller's view, at virtually all loci, there is a single optimal or "wild-type" allele. That the reference sequence is composed of the DNA of 24 chromosomes — the 22 autosomes as well as both X and Y sex chromosomes — and therefore (essentially) represents a haploid genome is also consistent with Muller's account. The assumption is that the normal or "wild-type" diploid genotype is made up of identical haplotypes with the optimal or "wild-type" allele present at each locus. Homozygosity is the norm — for all gene loci and the diploid genome as a whole. Heterozygous gene loci represent harmful deviations from the norm. This is consistent with Muller's emphases on partial dominance/incomplete recessiveness

and negative stabilizing selection and ignores the possibility that heterozygotes may exhibit equivalent or superior fitness compared with homozygotes — equivalent at loci where "deleterious" recessive alleles are completely recessive and superior at loci that are overdominant (heterotic). The HGP's adherence to a single functional genotypic norm also ignores, as did Muller, the possibility that populations may contain equally fit alternate genotypes that are supported by other forms of balancing selection such as frequency-dependent selection or diversifying selection. For Muller, as for the HGP, in the overwhelming majority of cases, variants are inferior deviates awaiting elimination by natural selection. The optimal or "wild-type" allele will almost always be the one that is found most frequently at each locus with the only exceptions being the small number of loci that are subject to heterotic or frequency-dependent selection and the very rare occasion upon which a new beneficial mutant allele is in the process of spreading through the population on its way to replacing its no longer optimal predecessor. The fitness of a population is maximized by the extent to which individual genotypes approach this single functional norm. Genetic variation is invariably harmful deviation or "genetic load."

Since Muller's time, we have become aware of the vast extent of genetic variation in natural populations. Consequently, it would be unsatisfactory to rest here in a comparison of the HGP's goal to produce a single DNA reference sequence with Muller's conception of a "normal" genome. However, the account can easily be updated. To retain the notion of a single functional norm that embodies genomic health, divergences from the reference sequence need either to represent differences that are wholly irrelevant to function or deviations that denote abnormal function or disease. This finds support in Kimura's neutral theory, particularly in its earliest formulation. Recall that Kimura sharply divided the genome into regions of two types: functionally constrained regions and functionally unconstrained regions. Functionally constrained regions are those of functional importance. These areas tend to be highly conserved in evolution and show little interspecific or intraspecific variation in nucleotide sequence because they are subject to negative stabilizing selection. Departures from the reference sequence in functionally constrained regions fall into two classes: "defects" that impair function and functionally equivalent differences, such as silent nucleotide base substitutions.

Functionally unimportant areas that are governed by drift exhibit a much greater degree of both interspecific and intraspecific variation. In these areas, the reference sequence loses its authority as a normative standard indicative of health and normal functioning. For molecular biologists, divergences from the reference sequence are "evolutionary noise"; the reference sequence is just one sequence among many. For evolutionary biologists and biological anthropologists, on the other hand, DNA sequences in such regions are helpful in reconstructing evolutionary histories. The human nuclear DNA reference sequence is likely to be partitioned for use into functionally more and less constrained regions as the human mtDNA reference sequence has been, where evolutionary biologists and biological anthropologists focus on variability in the less constrained "D-loop" region of the genome and molecular biologists interested in normal function, and clinicians interested in mitochondrial diseases, focus on protein-coding regions.

To draw this connection between Muller and Kimura may seem surprising: Muller emphasized the role of natural selection in evolution whereas Kimura's theory focuses on the importance of chance factors, at least at the molecular level. However, in his influential (1974) *The Genetic Basis of Evolutionary Change*, Lewontin refers to Kimura's theory as "neoclassical" in order to emphasize the historical and conceptual continuities between Muller's classical theory of the genetic structure of populations and Kimura's neutral theory.[21] The conceptual continuity lies in Muller's and Kimura's shared emphasis on negative "purifying" selection and their belief that natural selection acts to eliminate genetic variation and not to preserve it:

> the so-called neutral mutation theory is, in reality, the classical Darwin-Muller hypothesis about population structure and evolution, brought up-to-date. It asserts that when natural selection occurs it is almost always purifying, but that there is a class of subliminal mutations which are irrelevant to adaptation and natural selection. (1974a, p. 198)

[21] This chapter focuses on conceptual, as opposed to strict historical, continuity. See Dietrich (1994) for a challenge to "Lewontin's Historical Thesis."

Dobzhansky agreed with Lewontin's analysis:[22]

> The classical model is false. Its former partisans have made a clever about-face. The enormous amount of genetic variation now discovered in natural populations is biologically and adaptively insignificant. It is neither useful nor harmful; it is neutral. (1976, p. 102)

Kimura and Muller alike down-played the importance of balancing selection, both heterotic and nonheterotic. Intraspecific genetic variation is either selectively neutral/functionally irrelevant difference or harmful deviation that contributes to "genetic load." Genetic variation is virtually always in either mutation-drift or mutation-selection balance and only very rarely (and likely temporarily) contributes to fitness.

Thus far, I have argued that the HGP's goal to establish a single DNA reference sequence — representing a haploid human genome — as a standard of health and normal functioning is consistent with a set of evolutionary views shared by Muller and Kimura, at least with respect to functional regions of the genome. Both models predict that, in a population under selection pressure, little genetic variation will be found in functionally constrained areas of the genome. On this view, at least for functionally important regions of the genome, it would not seem unreasonable to expect that a composite DNA reference sequence compiled from a small number of presumably healthy individuals might provide a reasonable standard of health and normal functioning. These individuals, if typical members of the population, would carry normal or "wild-type" alleles, or their selectively neutral equivalents, at the vast majority of their loci. Likely, the most useful reference sequence to be used as a standard for comparison is one that includes, at each locus, the allele that occurs most frequently in the population. Infrequent neutral variants and deleterious mutants could be catalogued separately, as we have seen done for the mtDNA data base. Since the bulk of variation is expected to be neutral, improvements to the

---

[22] Dobzhansky believed, however, that Lewontin's designation of the "new panneutralist model" as "neoclassical" is misleading because of the different ways in which genetic variation is treated by the classical and neutralist positions:

> The prefix "neo" does not do justice to the basic difference between the classical and the panneutralist models. The keystone of the former was the assumption of the prevalence of genetic uniformity and of normal or wildtype chromosomes and genotypes. Panneutralists do not deny the prevalence of polymorphism and heterozygosity; they merely assume it to represent a kind of noise in the genetic system. (1976, p. 102)

reference sequence would involve, for the most part, the replacement of relatively rare polymorphisms with more frequently occurring alleles and, far less often, the routing out of deleterious mutant alleles. DNA sampled from an average individual is likely to result in a useful reference sequence with relatively few modifications necessary, the majority of which would be of the "bookkeeping" variety associated with including the most frequently occurring allele at a locus, insofar as most individuals in a population are believed to have normally functioning or "healthy" genomes. The (updated) classical account of the genetic structure of natural populations seems to furnish credibility to what, admittedly, appears to be an unlikely venture: the sequencing of parts of genomes from a small number of average or typical humans to compile a composite DNA reference sequence that is to be used as a species-wide genetic standard of normal functioning and health.

However, Muller would have been opposed to this conclusion. As outlined earlier in the chapter, he believed that the health of the human "gene pool" has long been deteriorating due to the effects of civilization and of increased mutation rates due to radiation exposure. He estimated that the average human falls at least twenty percent below the adaptive norm because of the effects of being heterozygous for at least eight, and "possibly scores," of deleterious alleles. A reference sequence compiled from the DNA of average individuals would not, therefore, provide an adequate genetic standard of normal functioning or health. For Muller, though, such a standard is theoretically possible, even if no actual genome conforms to it. Muller believed, in fact, that in the human evolutionary past, before the advent of civilization and its corrupting forces, the "normal" or optimal genotype prevailed among members of the species. This relates to the researcher mentioned in Chapter Two who suggests that the best mtDNA reference sequence would be that of "mitochondrial Eve" because her sequence would predate the species' accumulation of deleterious alleles. Evolutionary rationale for such a premise is found in the theories of Muller and Kimura. Muller considered the glory days for the human physique to have been those of the hunters and gatherers, in whom millennia of

harsh selective forces permitted nature to carve out a resilient genome.[23] On Kimura's "effectively neutral" model, since all mutations are at least slightly deleterious and approach absolute neutrality only in the limit, the species is constantly deteriorating ever so slightly. Kimura's "optimal" genome is older even than Muller's hunter-gatherer genome — it comprises the rare beneficial mutations from which the species was built. Hence, for Muller and Kimura alike, the "normal" genome is the "original" genome that is unblemished by mutation and prevailed in the distant evolutionary past. Since mutations are virtually always at least slightly deleterious, the "original" genome also represents a standard of normal or optimal functioning. Although many of the "original" and "optimal" alleles that contribute to this "original" and "optimal" genotype continue to exist in the population, the "healthy" genotype itself exists in no actual individual. Hence, it is an ideal in the senses both that it does not exist in reality and that it represents a normative standard to which all existing genotypes are inferior.

The commitment that Muller and Kimura share to the (theoretical) existence of a single optimal genotype follows from a set of basic theoretical assumptions about population structure that they hold in common. Both subscribe to models that attach affix intrinsic and fixed properties to individual alleles. A particular allele is "good," "bad," or "neutral" regardless of context: no matter what allele is present at the same locus on the opposite chromosome, no matter what alleles are present in the remainder of the genome, no matter what is the composition of alleles in the population, no matter what are the environmental conditions. This is exactly what was at issue in the "bean bag" genetics controversy. The "bean bag" label arose with Ernst Mayr in his 1959 Cold Spring Harbor address titled "Where Are We?" Mayr argued that the mathematical tradition in population genetics that is associated with Fisher, Wright, and Haldane wrongly attributes absolute selective values to individual genes and ignores that selective values are relative to genetic and environmental backgrounds in individuals and populations alike: "Evolutionary change was essentially presented as an input or output of genes, as the adding of certain beans to a beanbag and the withdrawing of others" (p.

---

[23] Muller believed that physical fitness is the product of natural selection but that mental fitness is the product of cultural evolution. The glory days for the latter lie ahead, therefore, and not behind.

2). It would be wrong to infer from Mayr's critique that mathematical approaches ignore the relativity of genic effects. Selective values for individual alleles are mean, not absolute, values. An allele's mean fitness is its average effect across the range of genetic and environmental backgrounds which it encounters. Mayr also overlooked the fact that Wright's mathematical models assume mean selective values for whole genotypes, not individual alleles, and thus incorporate the effects of genic interaction at the level of individual organisms (Provine 1986, p. 482).

However, for our present purposes, Mayr's dismissal of "bean bag" genetics raises several important points. The "good," "bad," or "neutral" allele is always found in *some* population with *some* kinds of genetic and environmental backgrounds, and Muller and Kimura share certain assumptions about these parameters from which Dobzhansky and other selectionists depart. For Muller, populations are large and randomly breeding; Kimura and Crow's (1964) "infinite alleles" model, upon which their introduction of the possibility of neutral alleles was based, assumes a finite but very large number of alleles in the population (Dietrich 1994, pp. 38-41). Muller and Kimura alike considered environments to be relatively constant across time and place and believed that individual organisms are protected from random environmental fluctuations by developmental homeostasis. Each assumed that the genetic health of the human species is gradually deteriorating over time, although for different reasons. Muller believed this to reflect a deteriorating environment — he considered the "harsh" and "natural" hunter-gatherer conditions to represent the ideal environment. Kimura, who assumed the constancy of environments over geological as well as generational time, conceived this gradual genetic deterioration of the species in terms of the departures in size of actual populations from the ideal, very large but finite, population of his mathematical models. As population size diminishes and the effects of drift become more pronounced, it is increasingly likely that deleterious mutant alleles will increase in relative frequency and eventually become fixed in the population.

Besides these idealizations of the "normal" or "optimal" genotype as the "original" species genotype, an alternate approach exists that attempts to improve the ability of a single reference sequence to serve as a suitable genetic standard of normal functioning and health: the replacement of a composite reference sequence with a consensus

reference sequence. In Chapter Two, we saw that this has been suggested with respect to the mtDNA reference sequence. Recall that a consensus sequence, at each nucleotide position, displays the nucleotide base that occurs most frequently among DNA sequences sampled from a number of individuals. The notion of a consensus sequence, at first glance, conveys a statistical sense of normality. Indeed, a recent patent application for a consensus sequence of the "normal" BRCA1 gene argues that this version is a superior reference sequence for the gene because it is likely to be the sequence that will be found most frequently in the population (Marshall 1997a). Aside from the fact that the sample on which the consensus sequence was based included only five individuals, and the unlikelihood that these individuals were of diverse ethnic origins, there is no guarantee that a sequence comprising the most frequently occurring nucleotides at each nucleotide position will itself occur most frequently in a given population, not to mention in the species as a whole. Nor is there any guarantee that a consensus reference sequence represents a functionally normal or healthy gene and is not just an idle statistical invention. To draw these conclusions — that is, that a consensus sequence represents both statistical and functional norms — additional evolutionary assumptions are required: specifically, that evolutionary forces operate at the level of individual nucleotides. This would mean, for example, that individual nucleotides are equally likely to be separated in recombination and that natural selection operates consistently at a given nucleotide regardless of genetic and environmental backgrounds.

Evolutionary grounds *do* exist for using a consensus reference sequence as a genetic standard of normal functioning and health. These are to be found in Kimura's "effectively neutral" version of neutralism in which he assumes that functional constraints operate continuously across the genome and that all mutations are at least slightly deleterious, approaching absolute neutrality only in the limit. If all individual nucleotide base substitutions, insertions, and deletions are at least slightly deleterious, even if these occur in regions of the genome that are under minimal functional constraint, it can be inferred that absolute fitness values attach to individual nucleotides and not just to individual alleles. Consequently, it makes as much evolutionary sense to refer to the "normal" or optimal nucleotide base at each of the approximately three billion nucleotide positions of the haploid genome as it does to speak of "normal" or optimal alleles. As

I have already argued, on the "effectively neutral" model, since all mutations are at least slightly deleterious and approach neutrality only in the limit, the "normal" or optimal genotype is the original genotype. Any consensus reference sequence that is compiled by sampling an actual population will fall short of this ideal to the extent to which the population has deviated in size at any time throughout its evolutionary history from the very large but finite population of Kimura's mathematical models.

Of course, the selectionist who emphasizes the evolutionary importance of gene-gene and gene-environment interactions, and the genetic composition of individual populations and the local environmental conditions that these populations experience, is hardly likely to grant fixed selective values to individual nucleotides that individual alleles do not themselves possess. For the selectionist, only individual genotypes that belong to an actual population with a particular genetic composition and a specific set of local environments can properly be assigned fitness coefficients. There is no ideal environment and there is no ideal population that determine absolute fitnesses of individual genotypes, alleles, or nucleotides. The selectionist emphasizes the "continuous changes" that occur in physical, biotic, and genetic environments (Mayr 1959, p. 6) and concludes that different genotypes are likely to prove adaptive in alternate sets of circumstances. Such fluctuations in genetic and environmental backgrounds are believed to contribute to the maintenance of genetic variation in a population through different forms of balancing selection. Consequently, on the balance/selectionist account, no single DNA reference sequence can represent a genetic standard of normal functioning and health — whether for the species as a whole or for individual populations. Divergences from the reference sequence need not be either irrelevant to function (because these are functionally equivalent or because they occur in nonfunctional regions of the genome) or abnormal deviation that is indicative of malfunction or disease. Multiple different modes of functioning may be similarly viable. Genetic variation need not constitute deviation from a single adaptive norm.

Any standard of normal functioning and health that the HGP's DNA reference sequence is purported to represent must be recognized to be contingent with respect to evolution in three ways: ontological, metaphysical, and epistemological. First, a standard of

genetic normality is evolutionarily contingent in an ontological sense. As Max Delbrück writes in one of the quotations that lead off the chapter, "there are no 'absolute phenomena' in biology. Everything is time-bound and space-bound" (in Mayr 1961, p. 1502). From the perspective of evolution, it is completely arbitrary to designate a single DNA sequence as representative of "the" human type. Such a representation can be no more than a snapshot taken at a single moment in the flow of evolutionary time; it is "time-bound." It is also "space-bound"; it ignores the local contingencies associated with the unique evolutionary histories of diverse human groups.

Second is the contingency associated with one's metaphysical stance toward change and diversity in the biological realm. As Georges Canguilhem notes, this stance depends "on whether one is a fixist [*fixiste*] or a transformatist" (1989, p. 141). We have seen that Muller and Kimura are "fixists" who treat "living beings [that] diverge from the specific type [as] abnormal in that they endanger the specific form" (ibid.) and emphasize the constancy of environments over time and space. Genetic variation is harmful deviation from an original optimum. Those who adhere to balance and selectionist positions, on the other hand, are "transformatists" who focus on the dynamic aspects of evolution and emphasize the diversity of environments over time and space. Genetic variation is beneficial to the species in the present as well as in the future.

Third, from an epistemological standpoint, distinct theoretical bases support these differing — "fixist" and "transformatist" — attitudes. The HGP's aim to produce a single DNA reference sequence that can serve as a standard of normal functioning and health finds a theoretical basis, as we have seen, in the classical and neutralist accounts of the genetic structure of populations. Although the classical/neutralist and balance/selectionist positions understand genetic variation, genetic mutation, and the "normal" genome in such divergent ways, the empirical evidence has been inadequate to decide between them. As Crow (1981) notes, the distribution of genetic polymorphisms found in population studies is consistent with both hypotheses. In addition, experiments to determine selective neutrality can only measure selective differences to within three to five percent which, according to Crow, is far from neutral in a large population. Therefore, there is no way to discriminate between hypotheses of neutrality and of weak selection. The controversy is still ongoing; as one population geneticist recently writes: "Whether the amount of

genetic variation in a population is maintained by natural selection or by random genetic drift of neutral mutants is one of the most important issues in population genetics" (Tajima 1997, p. 149). Insofar as the extent to which genetic variation is selected *for* or selected *against* remains unknown, generally and for specific regions of the genome, any standard of genetic normality that the HGP's DNA reference sequence is purported to represent must be recognized to be evolutionarily contingent from an epistemological, as well as metaphysical and ontological, perspective. This is something quite different, though, from regarding the HGP's mapping and sequencing goals to be "anti-evolutionary."

3.3.2 Typological Thinking and the Human Genome Project

As I outlined in the introduction to this chapter, evolutionary biologists and philosophers of biology have criticized the thinking that informs the HGP as pre-Darwinian, Platonic, essentialist, and typological for its treatment of genetic variation as deviation from a norm and mutation as "error" or "damage." Since I have argued that the HGP's aims are consistent with classical and neutralist accounts of the genetic structure of natural populations, it is not surprising that Dobzhansky similarly accused Muller of pre-Darwinian, Platonic, essentialist, and typological thinking for his adherence to the notion of a "normal" genome. Nor is it surprising that Lewontin refers to Kimura's neutral theory as "neoclassical": "the so-called neutral mutation theory is, in reality, the classical Darwin-Muller hypothesis about population structure and evolution, brought up-to-date" (1974a, p. 198). In this final part of the chapter, I attempt to delineate exactly what charges of pre-Darwinian, Platonic, essentialist, and typological thinking involve. I assess possible ways in which the HGP may be guilty of such a charge with reference to Dobzhansky's analogous criticisms of Muller. I conclude by considering possible reasons why molecular biology's implicit evolutionary assumptions might coincide with Muller's account.

The typological-population distinction dates to Dobzhansky and Mayr. The distinction is said to have been first fully articulated by Mayr in 1959 (Sober 1984b, p. 14). It centres on how biological variation is conceptualized. Typological thinking assumes that it makes sense to speak of a "type" that is representative of an entire class

of organisms — whether at the level of genus, species, or race. On this view, there are definite, and more or less fixed, qualitative differences that distinguish one genus, species, or race from another. Variation within a genus, species, or race is understood as deviation from type. Population thinking refuses this notion of a normal type. Various genuses, species, and races are recognized to differ in degree and not in kind. The boundaries that separate them are understood to be contingent products of evolution in a state of ongoing flux. Intragroup variation is understood simply as difference and not as deviation from type.

Much rhetoric surrounds Dobzhansky's and Mayr's promotion of population thinking in their writings. Typological thinking is described as pre-Darwinian, static, essentialist, Platonic, *a priori*, morphological, prescriptive, and idealist. Population thinking is described as Darwinian, dynamic, individualistic, statistical, empirical, and genic. The following passage is typical of Dobzhansky's criticisms of Muller's adherence to the concept of a "normal" genotype and his classical theory of the genetic structure of populations as typological:

> It is legitimate to use the concept of "norm" to facilitate the description of mutants and of genetical and environmentally induced aberrations that occur from time to time. Unfortunately, some biologists have gone beyond this, and came to regard the "norm" as a sort of ideal prototype of which the actually existing individuals are imperfect copies. This typological thinking, the roots of which go down to the Platonic philosophy, is basically anti-evolutionistic, and has produced much confusion in biological thought. (Cordeiro and Dobzhansky 1954, p. 83)

To assess the validity of Dobzhansky's criticisms of Muller, it is necessary to try to figure out the meanings of some of these descriptions.

Mayr argues that, unlike populationists, typologists fail to appreciate that natural selection is a statistical phenomenon:

> The typologist interprets natural selection as an all-or-none phenomenon. He assumes that one type is better and therefore survives, while the other type is inferior and is therefore wiped out. Natural selection in this interpretation is immediate, absolute, and final. (Mayr 1963, p. 184)

Typologists, in addition, emphasize selection's negative effects: "Natural selection.... either selects or rejects, with rejection being by far more obvious and conspicuous" (Mayr 1984, p. 17). An expectation for "every population to consist uniformly of perfect

individuals" (Mayr 1963, p. 184) results. The populationist, on the other hand, "does not interpret natural selection as an all-or-none phenomenon" (Mayr 1984, p. 17):

> Every individual has thousands of traits in which it may be under a given set of conditions selectively superior or inferior in comparison with the mean of the population. The greater the number of superior traits an individual has, the greater the probability that it will not only survive but also reproduce. But this is merely a probability. (ibid.)

The population approach emphasizes the importance of variable environmental conditions in natural selection and shifts attention away from the relative fitnesses of different morphological "types" to an operational understanding of fitness in terms of probable changes in gene frequencies from one generation to the next. Mayr considers typological thinking to be "pre-Darwinian" not only because it ignores the statistical treatment of natural populations presented in Darwin's theory of evolution by natural selection but because it is "essentialist": species are conceived of in terms of fixed unchanging essences with "complete discontinuities" between all types (1963, p. 5). Darwin's conception of evolution by natural selection as a gradual process is completely incongruous with species essentialism. Mayr reaches the conclusion that the typologist who is also an evolutionist must believe in saltatory evolution: "Since there is no gradation between types, gradual evolution is basically a logical impossibility for the typologist. Evolution, if it occurs at all, has to proceed in steps or jumps" (1984, p. 15).

Other characterizations of the typological-population distinction found in Mayr and Dobzhansky centre on the treatment of variation within a population, race, or species. Plato's theory of the forms is appealed to in order to illustrate the differences between typological and population thinking: "Plato's concept of the *eidos* is the formal philosophical codification of [typological] thinking" (Mayr 1963, p. 5); typological thinking "regard[s] the 'norm' as a sort of ideal prototype of which the actually existing individuals are imperfect copies" (Cordeiro and Dobzhansky 1954, p. 83). Mayr and Dobzhansky emphasize that, for populationists like themselves, definitions of species and subspecies categories are not real but only ideas. Rather, it is the observed variability of nature that continually defies containment in these categories that is real. Plato, of course, believed that empirical properties only imperfectly approach the true, eternal, and unchanging ideal Forms. In Dobzhansky's words: "Plato was not a biologist ... but

following him we would have to conclude that the mice or the flies we catch are only imperfect copies of the one ideal Mouse or the ideal Fly" (1962, p. 42). In contrast to this tie between typological thinking and philosophical Platonism, Dobzhansky held that population thinking shares philosophical existentialism's focus on the importance of individuality: "the populational approach ... considers the differences among people and the variations among individual animals and plants to be very real and important, not just appearances or accidents or imperfections" (ibid., p. 43). Mayr similarly points to the stress population thinking places on "the uniqueness of everything in the organic world": not only are no two individuals alike but each individual changes throughout its life cycle and when confronted with different environments (1984, p. 15).

Although Muller did emphasize negative "purifying" selection, he cannot be considered a typologist of the sort Mayr describes as a "pre-Darwinian" and "essentialist" adherent of saltatory evolution or evolution-sceptic. Dobzhansky's claim that Muller's thinking is "basically anti-evolutionistic" is entirely unfair. Muller was a tireless advocate of neo-Darwinism who, even as a student, tangled with T. H. Morgan over Morgan's support for saltatory evolution and mutation as the sole basis for evolutionary change. To accuse Muller of "Platonism" is not entirely fair either. Muller did not believe that there is "one ideal, perfect, and inconceivably beautiful man" of which the "[p]eople whom we meet are ... only more or less defective and distorted images" as Dobzhansky's caricature suggests (1962, p. 42). Muller's "ideal type" is the product of Nature; it does not reside in Plato's heaven. Muller also stressed that, because natural selection lacks foresight, "[t]he organism is not perfect in any absolute sense" but "relative only to the possibilities more immediately around it."

Where Dobzhansky's criticisms of Muller ring true are in Muller's conception of the normal. The typological-population distinction that opposes "types" to "individuals" represents differing conceptions of biological norms. As Mayr writes: "For the typologist everything in nature is either "good" or "bad," "useful" or "detrimental" (1984, pp. 16-17). In other words, biological variation is always deviation from an adaptive norm. Populationists, on the other hand, prefer statistical norms: "All organisms and organic phenomena are composed of unique features and can be described collectively only in

statistical terms" (ibid., p. 15), that is, in terms of the arithmetic mean and variance for the population:

> Averages are merely statistical abstractions; only the individuals of which the populations are composed have reality.... For the typologist, the type (*eidos*) is real and the variation is an illusion, while for the populationist the type (average) is an abstraction and only the variation is real. No two ways of looking at nature could be more different. (ibid., p. 16)

The typological-population distinction written this way is one that we are familiar with from Chapter Two's comparison of the statistical approaches of Adolphe Quetelet and Francis Galton.

Galton understood the mean values of traits to be, in Mayr's words, "statistical abstractions." For Quetelet, on the other hand, the constancies in mean values of traits reflect the operation of biological (or social) laws or "constant causes" whereas variation follows from "accidental causes." Mean values are "ideal types," variation from which constitutes harmful deviation. This is the basis of Elliott Sober's understanding of the distinction between essentialist (typological) and population thinking in his (1980) article, "Evolution, Population Thinking, and Essentialism." Sober characterizes essentialism in terms of the approach to biological variation that was taken by Aristotle and Quetelet. Aristotle's "Natural State Model" holds that "there is a distinction between the *natural state* of a kind of object and those states which are not natural. These latter are produced by subjecting the object to an *interfering force*" (p. 360). Variability in nature is thus to be understood as "deviation from what is natural" (ibid.). Similarly, for Quetelet, "variation in a population ... is the result of interferences confounding the expression of a prototype" (p. 367). Sober argues that this conception of variation as deviation became no longer tenable with the rise of statistical population thinking in the late-nineteenth century in the work of Galton and Darwin.

Yet, just as Quetelet opposed constant to accidental causes, Muller, a committed follower of Darwin, opposed deterministic to random forces of evolution — negative stabilizing selection to mutation and drift. His conception of a single genotypic norm that is homozygous for the "wild-type" allele at all loci is analogous to Quetelet's "ideal type" from which variation represents harmful deviation. Muller's "normal" genotype is vulnerable to the same criticisms faced by Quetelet's "average man": the "ideal type"

does not exist. Dobzhansky charged that Muller's "ideal Man — or man-as-he-might-have been" has never existed (1959, p. 158) and Mayr writes: "An individual that will show in all of its characters the precise mean value for the population as a whole does not exist. In other words, the ideal type does not exist" (1984, p. 16). Yet, neither Muller nor Quetelet was misled on this count: Muller considered his "all-normal man" to be "fictitious" and Quetelet viewed his "average man" as "an identity [that] can scarcely be realised." The important issue, I would argue, is not whether the "ideal type" actually exists but what it is understood to represent *as* an ideal. For Muller, the "ideal type" *is* somewhat like Plato's forms because, although it cannot be perfectly attained, just as perfect knowledge of the forms is impossible, analogously, through reason, it is possible to come close. Humans can learn to control mutagenesis and to direct evolution. By decreasing the spontaneous mutation rate and by instituting negative eugenics, it may be possible to restore, at least in part, the optima for physical traits that belong to our "noble savage" past. Mental perfection remains an ideal to strive for in the future. That Muller's "ideal type" does not, and even may never, exist in reality does not, by itself, impugn its status as a norm.

As with Muller, the charge of "typological" thinking sticks to the HGP in some of its aspects and not in others. The HGP is not anti-evolutionary or pre-Darwinian; rather, the HGP's aim to produce a single DNA reference sequence that represents a genetic standard of normal functioning and health is consistent with a particular set of evolutionary assumptions — elements of which are shared by Muller and Kimura. On the other hand, like Muller, the HGP exhibits a typological conception of genetic variation as deviation from an adaptive norm or "genetic load." Elof Axel Carlson, Muller's former student and biographer, contends that "[Muller's] theoretical contributions to genetics and evolution provided, in large measure, the world-view which molecular biology has adopted" (1973, p. vii). While I believe that Carlson is right that molecular biology shares Muller's evolutionary world view, the reasons for this are not transparent. Perhaps, as well, it would be more accurate to say that evolutionary assumptions like Muller's are implicit in the theories and practices of molecular biology. Certainly, the Muller-Dobzhansky debate had little impact on molecular biology during the 1950s and 1960s. The controversies over genetic load and radiation dangers had very much to do with the

genetics of human populations, and, while they drew on the extensive data on *Drosophila* that was available, these issues were remote from the genetics of the microorganisms that molecular biologists studied. By this time, molecular biology was well off and running under its own steam, with its own set of problems.

Nevertheless, there are several other possible reasons for this appearance of conceptual continuity. The first obvious one is that most biologists, other than naturalists, hold evolutionary views similar to Muller's: adaptation as directional, mutations as good, bad, or neutral regardless of context. Certainly, the scant references to evolution and mutation found in the first molecular biology texts (written in the 1960s), as well as the more extensive writings on evolution found in molecular biologists' more popular works, express views similar to Muller's.[24] However, whereas Muller's beliefs about evolution underwrote his conceptions of normality and variation, it is likely that questions about evolution and populations are, and can be, completely ignored by molecular biologists. Their questions about structure and function couched in terms of the normal and abnormal are not informed by, but merely consistent with, an evolutionary scheme that is similar to Muller's.

Another reason why the views of Muller and molecular biologists coincide is their shared indebtedness to the concepts and mechanisms of "classical" genetics. It seems that Dobzhansky fixed the "classical" label to Muller's position in reference to its tie to the mutant-wild type distinction, in writing that his own balance position renounces the validity of the "[t]he classical concept of the 'wild-type,' and the distinction between 'normal' alleles ... and mutant alleles, [which] arose in *Drosophila* genetics" (1955a, p. 4). There is no doubt that the bridge biochemical and bacterial genetics provided from "classical" genetics to molecular biology preserved and reinforced the mutant-wild type

---

[24]These are, in short: Due to evolution by natural selection, organisms are extremely well-adapted to their environments and function efficiently within them, the products of the accumulation over millions of generations of very infrequently occurring fitness-enhancing mutations. Hence, the vast majority of mutations that occur are detrimental to the organism and involve the loss or impairment of protein function. These are eliminated from the population by natural selection. The only mutations that will persist in the population are those rare ones that are neutral (such as eye colour in humans) or, rarer yet, beneficial with respect to protein and organismal function. Where a mutant allele confers an advantage, natural selection rapidly leads to the elimination of its predecessor from the population.

distinction, and provided a chemical basis to several of the key concepts in "classical" genetics that support Muller's evolutionary views. The one gene-one enzyme hypothesis of George Beadle and Edward Tatum, and its subsequent incarnation as the one gene-one polypeptide model, as well as the definition of the gene as a functional unit (a cistron) as opposed to a unit of recombination or mutation, reinforce the notion of the "wild type" as functional norm. The normal or wild type gene produces a functional enzyme or polypeptide; the mutant gene produces a functionally deficient enzyme or polypeptide. "Forward" and "backward" mutations are identified on the basis of whether some biochemical function is lost or restored (Freese 1963, p. 210). Thus, the "classical" concepts of dominance and recessiveness and the laboratory experiences of "classical" geneticists that the vast majority of mutations are recessive and involve loss of function (vestigial wings, for example) are understood in biochemical terms. Muller's belief in the ubiquity of partial dominance is vindicated: the heterozygote appears "wild type" only if there are sufficient amounts of normal protein present.

The normal-abnormal distinction is central to molecular biology, not just because it arose specifically out of the laboratory approaches of classical, biochemical, and bacterial genetics, but because these fields alike share the interventionist approach of experimental biology. Abnormalities are induced in order to study normal function. There are two features of this approach consistent with that of Muller to evolution. The first is that the existence of the normal organism, trait, or gene, is assumed from the outset. The statuses of the mutant and normal gene are never doubted in experimental genetics because mutations can be discerned *as* sudden heritable changes in a characteristic, whether the modification is perceived at the level of a gross phenotypic characteristic, a gene product, or a nucleotide base. The "normal" or "wild type" organism/characteristic/gene/allele is uncontested: it is that which was already present. Mutation is a tool in the laboratory; its responsibility for generating diversity in nature can easily be ignored. The second consequence is that experimenters must control for all other possible confounding variables. One inbreeds or clones organisms and maintains the external environment constant. Hence, just as for Muller, there is no interacting system; context is denied, the organism and its environment are effaced.

By maintaining genetic and environmental backgrounds constant in the laboratory, it is possible to attribute to individual alleles absolute selective values and constant functional effects. In this way, the laboratory geneticist shares the "bean bag" approach to population genetics that was parodied by Mayr (1959). Indeed, Wallace (1991) associates the "classical" label for Muller's views on the genetic structure of populations with the mathematical evolutionary genetics of Fisher and Haldane, rather than with the mutant-wild type distinction of *Drosophila* genetics as suggested by Dobzhansky's writings. This presents a challenge to Sober's (1980) argument that the essentialist and typological approach to biological variation found in Aristotle and Quetelet "has been discredited by modern evolutionary theory" (p. 365). Sober concludes that, although "our 'modern' conceptions of health and disease and our notion of normality as something other than a statistical average enshrine Aristotle's model" (p. 363), this cannot be justified by modern evolutionary theory. In other words, essentialism is a problem that persists in medicine and molecular biology but not in population genetics. In a footnote, Sober remarks that he sees no basis for Mayr's assertion that "essentialist errors continue to be made in population biology in the form of the distortions of 'bean-bag genetics'" (p. 353). However, as I have argued in this chapter, the population approaches of Muller and Kimura take typological and essentialist approaches to genetic normality and genetic variation that assume that individual alleles have intrinsic and fixed selective values and that there is an "original" and "optimal" genotype. Supposing that the fitnesses of alleles are independent of their genetic and environmental contexts and that the effects of alleles are strictly additive, the "normal" organism becomes conceived as the "bean bag" bearer of all normal, and no mutant, alleles.

3.4 Summary

In this chapter, I have argued that molecular biology and the HGP are not anti-evolutionary. Rather, their approaches are consistent with a particular set of evolutionary assumptions. In correspondence with Kimura's early formulation of the neutral theory, the genome tends to be divided into functional and nonfunctional regions. Muller's

classical account, as updated by Kimura's initial version of the neutral theory, permits at least functional areas of the genome to be represented by a single functional norm. This means that all genetic variation in functional areas of the genome is suspect. As updated by Kimura's later "effectively neutral" model, the classical-neutralist account provides evolutionary justification for the entire genome to be represented by a single functional norm. This means that all genetic variation is suspect. Since all mutations are deleterious — even if just slightly — the "normal" or optimal genome is the original genome, with very few exceptions. Individual nucleotides, as well as individual alleles, are assumed to have constant selective values: this justifies the notion of a consensus DNA reference sequence. To the extent to which the reference sequence approximates the "normal" or original genome, variation from it can rightly be judged to be harmful deviation. Although this sequence may belong to no *actual* individual, and is ideal in this sense, its normative status finds theoretical justification in the accounts of Muller and Kimura.

Although there are no obvious historical reasons that explain why molecular biology incorporates evolutionary assumptions similar to those of Muller, I offered several other possible reasons for this coincident set of operational precepts. I argued, also, that Dobzhansky's criticisms of Muller's approach as "typological" are warranted with respect to Muller's treatment of intraspecific genetic variation as deviation from a single adaptive norm. They are *not* warranted because Muller assumes that the "ideal type" actually exists or because his views are anti-evolutionary. Insofar as the approaches of molecular biology and the HGP are consistent with evolutionary assumptions found in Muller and Kimura, the case concerning current criticisms of the HGP as "typological" by evolutionary biologists and philosophers of biology is analogous.

Defenders of "bean bag" genetics have convincingly responded to Mayr by arguing that the selection coefficients of individual alleles are not absolute values that ignore genetic and environmental contexts; rather, they are mean values that take full account of the range of genetic and environmental backgrounds experienced by particular alleles and therefore any gene-gene and gene-environment interactions that occur. However, as Dobzhansky and other selectionists emphasize, selection coefficients will vary greatly in smaller populations with idiosyncratic genetic compositions and specific local environments. Among individuals, there is even greater potential for variable allelic

effects due to gene-gene and gene-environment interactions to arise. The legitimacy of a DNA reference sequence as a standard of normal functioning and health against which individual DNA sequences can be judged cannot rely for its justification on the mean fitnesses appealed to by defenders of "bean bag" genetics. Only if the selective values of individual alleles or individual nucleotides *are* constant and additive can one infer reliably from their behaviours in a population to their behaviours in a given individual. Insofar as human molecular genetics is concerned with the development of traits in individuals and molecular medicine is concerned with interventions at the level of individuals and not the "gene pool," the physiological effects of variation in individual alleles or nucleotides cannot be "black-boxed." The following chapter focuses on how we are to understand the value of genetic variation with respect to the physiologies of individual organisms and the clinical context.

Chapter Four

The Clinical Context: Is Genetic Variation Deviation or Deviance?

> It seemed a very small toe to cause such a degree of anxiety. But there was often a great deal of grown-up fuss that seemed disproportionate to causes. (Wyndham 1958, p. 13)
>
> A map of the human genome could ... lead to a more narrowly focused view of a "normal" gene complement, and how much deviation we permit before considering any individual genome "abnormal," deviant, or diseased. We haven't seriously begun to think about how to think about this issue, even though we know normalcy will be invented, not discovered. (Annas 1989, pp. 20-21)
>
> There are many ways to represent the nature of human beings, and none of them are [*sic*] value neutral. Even a genomic characterization is already always determined by our social and conceptual background. What we see, therefore, in a genomic characterization of human beings depends on what we are accustomed to and interested in seeing, this for both the species as a whole and the individual in particular. There is no escaping this immersion in the social and conceptual preconditions of observation, representation, science, and language; we cannot ever hope to achieve the position of an entirely unconditioned, uninterested observer. (Murphy 1994, p. 7)
>
> They wrestled with the novel idea that a Deviation might not be disgusting and evil — not very successfully. (Wyndham 1958, p. 53)

In this chapter, I further consider the normative status of the concepts of genetic normality and genetic mutation. Chapter Two illustrated ways in which several concepts of normality found in contemporary molecular genetics — wild-type, consensus sequence, and reference sequence — incorporate both descriptive and evaluative senses of the normal and often conflate them in confusing and misleading ways. Philosophical analysis may help to discern and to clarify these senses but it cannot, on logical grounds, proscribe the slippage that occurs between them. This is because, as Chapter Three showed, the biologist expects that due to natural selection what is statistically normal (frequent, common, average, or usual) will coincide with what is biologically (functionally) normal. These two senses of 'normal' — the statistical and the biological — are alike in that both are considered to be scientifically objective. By scientifically objective, I mean that judgements of whether particular genes or genomes are normal or abnormal are not informed by moral, aesthetic, social, or cultural values. Statistical judgements of genetic

normality are easily recognized to be objective and value-neutral: normal genes are those that occur most frequently; mutant genes are unusual variants. Judgements of normal and abnormal function are arguably evaluative but remain objectively scientific if they incorporate only biological values — survival and reproduction, for example.

The scientific objectivity of the concepts of genetic normality and genetic mutation is a matter of considerable interest and importance because knowledge in human molecular genetics is increasingly used to justify clinical (and social) interventions. For many theorists, the distinction between healthy or normal genes and disease or mutant genes constitutes a directive to action that resolves eugenic concerns surrounding the Human Genome Project and the increasing availability of genetic technologies. The purportedly scientific and objective line between genetic normality and genetic mutation underwrites boundaries between health and disease, enhancement and corrective therapy, and positive and negative eugenics. This permits modifications of individual genomes, at least on one side of the line, to be referred to as "therapy" or "intervention," expressions which accept the reassuring caress of the doctor's healing hand and displace the manipulative hand of the experimentalist.[1] Prenatal screening (with *in vitro* fertilization and embryo selection or selective abortion) and genetic manipulation (whether somatic or germ-line) are acceptable if they aim only to restore, but not to enhance, normal function. Ethicists write: "The object of germ-line therapy should ... be to restore an 'original' healthy genetic topology to the treated individual, such that future procreation would proceed as if one's progenitors had never carried a genetic lesion" (Zimmerman 1991, p. 599) and that germ-line intervention with "*bona fide* therapeutic purpose ... merely aims at restoring an order of things that obtained previously, but was disturbed by genetic mutation" (Mauron and Thévoz 1991, p. 656).

These appeals to a genetic norm to justify genetic manipulations of human germ cells are consistent with functionalist accounts of health and disease that define the

---

[1] Hubbard and Wald (1993, p. 110) use the term 'gene manipulation' instead of 'gene therapy' because the latter takes it for granted that the intervention is in the recipient's interests insofar as it seeks to restore health, a positive state. Since it is the very statuses of the distinctions between health and disease and genetic normality and genetic mutation that are in question here, I have chosen to follow them.

concepts of health and disease in terms of normal and abnormal biological functions. The philosophical project that provides clinical medicine with theoretical foundations in biology is an attempt to justify interventions designed to prevent and to treat disease. The "ought" of clinical intervention rests on the "is" of biological fact. If health is normal function and disease is abnormal function, the restoration of health and the eradication of disease represent no more than the preservation of what is "natural." Although functional accounts of health and disease have traditionally been based in physiology, they can easily be extended to the level of the genome with health and disease defined in terms of normal and abnormal gene structure and function. However, if George J. Annas and Timothy F. Murphy are correct in their views expressed in the quotations that lead off the chapter that "normalcy will be invented, not discovered" (Annas 1989, p. 21) and that "[e]ven a genomic characterization [of human nature] is already always determined by our social and conceptual background" and is not value-neutral (Murphy 1994, p. 7), there is no "is" of biological fact that does not also represent moral, aesthetic, social, or cultural "oughts."

The first section of the chapter discusses different possible accounts of health and disease based in normal and abnormal biological functions. I begin by introducing the best known of these: Christopher Boorse's (1977) functionalist account of health and disease. I discuss Boorse's goal-centred conception of functions in terms of two more recent philosophical approaches to biological functions: dispositional (or "forward-looking") and etiological (or "backward-looking"). Functionalist accounts of health and disease supported by either of these theories of biological functions make two key assumptions. The first is that biology provides theoretical foundations for clinical practice. The second is that judgements of normal and abnormal biological function, and those of health and disease that they support, are scientifically objective and not influenced by moral, aesthetic, social, or cultural values. The middle sections of the chapter deal with the first assumption: that theory directs action. I present Georges Canguilhem's two-part thesis in *The Normal and the Pathological*: first, that knowledge of normal physiology is indebted to a prior knowledge of the pathological, and not vice versa; and second, that knowledge of the pathological is directed by clinical judgements of health and disease, and not vice versa. Taking a close look at the methodologies of human molecular genetics, specifically the mapping of normal and mutant genes, I argue

for the extension of Canguilhem's thesis from physiology to human molecular genetics. In the final part of the chapter, I turn to the second assumption: the scientific objectivity of judgements of normal and abnormal biological functions and health and disease. If knowledge of the pathological is antecedent to and constitutive of knowledge of the normal, any standard of genetic normality will be no more or no less value-laden than judgements about what counts as a disease. I argue that our understandings of health and disease are always socially and culturally situated and, insofar as these are prior to and underlie biological designations of genetic normality and genetic mutation, the meanings of individual DNA sequences are no less socially and culturally embedded. This should not lead us, however, to throw up our hands and to move too quickly away from what has been at issue in the long-standing debate in the philosophy of medicine over the concepts of health and disease. Functionalist conceptions of health and disease teach us something important: that functions are always relative to an environment. Human environments are social and cultural, as well as biological. Appealing to several examples of pseudo-hermaphroditic conditions, I argue for an account of health and disease and genetic normality and mutation that recognizes the cultural negotiability of their meanings.

4.1 Functionalist Accounts of Health and Disease

Christopher Boorse bases his influential (1977) functionalist account of health and disease in the intuition that "the normal is the natural" (p. 554). Boorse establishes a theoretical medicine that defines health and disease in terms of normal and abnormal biological functions. Insofar as health and disease receive theoretical definitions based in empirical fact, theoretical justification is provided for practical interventions that aim to eliminate disease and to restore what is "natural." For Boorse, "[f]unctions are, purely and simply, contributions to goals. Any goal pursued or intended by a goal-directed system may serve to generate a function statement" (1984, p. 376). Since organisms are mechanistic goal-directed systems, a biological function is a *causal* contribution to a goal. Boorse argues that the goals attributed to goal-directed systems, and therefore the functions accorded to their component parts and processes, involve pragmatic choices dictated by the contextual

features of enquiry. Physiologists, evolutionists, and ecologists who are concerned with different goal-directed systems — organisms, populations or species, or ecosystems — will assign functions to different entities in nature.

Boorse follows a long tradition that takes physiology to be the appropriate theoretical foundation for a scientific medicine. This choice fixes the contextual features of enquiry: "In physiology the goal-directed system *S* is the individual organism and the relevant goals its own survival and reproduction" (Boorse 1984, p. 383). Physiological functions are assigned to whatever parts and processes contribute "reliably" to the goals of survival and reproduction "throughout a species or other reference class" (ibid.). Members of a given species share a "uniform functional organization" constituted by a "means-end hierarchy" of functions that Boorse calls the "species biological design" and which can be discerned empirically (Boorse 1977, p. 557). The theoretical concept of disease as deviation from "species biological design" (ibid., p. 543) includes not only those conditions we generally consider to be diseases but congenital defects, functional losses, and injuries. Boorse deems functions to be normal or abnormal in the statistical sense: a particular part or process functions normally if it makes a "typical contribution" to "individual survival and reproduction ... within members of a reference group" (ibid., p. 562):

> *Health* in a member of the reference class is *normal functional ability*: the readiness of each internal part to perform all its normal functions on typical occasions with at least typical efficiency.
> A *disease* is a type of internal state which impairs health, i.e., reduces one or more functional abilities below typical efficiency. (ibid.)

Boorse considers these theoretical definitions of health and disease to be objective and value-neutral because they are empirically based.

Boorse's basic account of biological functions is similar in several respects to that of Robert Cummins. Cummins refers to capacities rather than goals. Functions are identified by analyzing the capacities of a system in terms of the contributing capacities of its component systems and parts. All functions are contained within a more complex integrated whole (a "containing system") and can be isolated only relative to that whole, and, importantly, how one *chooses* to delineate and to analyze the system. Boorse and Cummins also agree that "functional analysis can properly be carried on in biology quite

independently of evolutionary considerations" (Cummins 1984, p. 399). Suppose, Cummins says, that pigeons' wings no longer contribute, and even become detrimental, to their fitness. We would still analyze pigeons' capacity for flight in terms of the functions of their anatomical parts. Boorse would likely argue that the anatomist may very well do so, but the physiologist, who understands the "apex goals" of the organism to be survival and reproduction, would not. Despite his focus on components of evolutionary fitness (survival and reproduction), Boorse nevertheless contends that physiologists need only be concerned with how a system operates at present and not with how it evolved. Although a trait's evolutionary origins may be helpful in gaining an understanding of its current function, the concept of biological function does not itself require this.

Recent philosophical theories of functions, function-ascribing statements, and functional explanation are more emphatically realist. Both "forward-looking" dispositional and "backward-looking" etiological accounts consider functions to be natural properties that exist independently of particular conceptual schemes and can be explicated by the theory of evolution by natural selection. Where these two approaches diverge is over the relationships of functions to evolutionary processes. The "forward-looking" approach represents all biological functions as propensities for survival and reproduction. On this view, functional explanations address "the adaptive significance of a trait observed in present individuals in a given environment" (Horan 1989, p. 135). Functions are adaptive but are not necessarily adaptations, that is, the products of past evolution by natural selection. The "backward-looking" approach is indebted to Larry Wright's thesis that functional ascriptions are explanatory only insofar as functions are identified on the basis of the evolutionary *history* of the particular part or process. Natural selection answers the question, Wright says, of "how the thing with the function *got there*" (1984, p. 359): "the *function* of the liver is that *particular* thing it is good for which explains why animals have them" (p. 359). Functions are adaptations; in present environments, they may or may not be adaptive.

John Bigelow and Robert Pargetter (1987) are responsible for a "forward-looking" or dispositional theory of functions. "Something has a (biological) function," they argue, "just when it confers a survival-enhancing propensity on a creature that possesses it" (p.

192). Functions are dispositional properties of biological entities: where natural selection operates, given particular environmental conditions, functions increase the likelihood of their bearers' survival. Bigelow and Pargetter stipulate that functions are relative to a creature's "natural habitat" although they admit this to be ambiguous in the event of a sudden change in the environment. Functions are assigned to components of organisms based on their contributions to the functioning of subsystems that are hierarchically arranged. "Habitat" extends to internal environments — organelles of a cell, for example, insofar as they contribute ultimately to organismal survival and reproduction, possess functions relative to the intracellular environment. Bigelow and Pargetter believe that their account intersects with that of Boorse, except in their appeal to propensities where Boorse appeals to "statistically normal activities within a class of organisms" (footnote, p. 193). Since my concern in this chapter is with functional accounts of health and disease, insofar as Boorse defines health and disease in terms of *physiological* functions and these, in turn, are defined by their contributions to the abilities of individual organisms to survive and to reproduce, I believe that it is appropriate to include Boorse among those theorists who take a "forward-looking" approach to functions based in ongoing evolutionary success.

The "backward-looking" or etiological approach asks not *how* a part or process works but *why* that particular part or process exists at all.[2] Along these lines, Ruth Garrett Millikan (1989a,b) and Karen Neander (1991) propose etiological accounts that emphasize "selective" or "proper functions": "biological proper functions are effects for which traits were selected by natural selection" (Neander, p. 168). Millikan (1989b) argues that "forward-looking" treatments of functions make improper appeals to natural selection because selection can explain a trait's presence (its origin or maintenance) in a population or species only if it is "temporally prior." "Forward-looking" accounts cannot handle traits that are universal to a species because a particular trait contributes to fitness only relative to other possibilities. Millikan and Neander believe Boorse's statistical determinations of normal and abnormal functions to be misguided. Biological norms are

---

[2] Mayr's (1961) paper distinguishes "how" and "why" questions according to analogous distinctions between proximate and ultimate causation and functional and evolutionary biology.

not statistical. A "proper" or "selected" function of a trait may be rarely fulfilled in actuality. "The notion of a 'proper function' is the notion of what a part is *supposed* to do" (Neander 1991, p. 180) — that is, "whatever it was selected for by natural selection" (ibid., p. 183) — and not what it *actually does* do. The selective history of a part or process determines its "proper function" and establishes biological (functional) norms. We identify a part or process as "defective" only through knowledge of the "proper function" for which it was favoured by natural selection in the past. The environment in which the organism finds itself is identified as "normal" only if it coincides with that in which the "proper function" was selected for in the past (Millikan 1989a, p. 300). It is because of natural selection and the biological/functional norms it establishes that objective judgements about health and disease are possible at all.

These functionalist accounts of health and disease, whether "forward-" or "backward-looking," can easily be extended to the level of the genome to found a "genetic medicine." Whether the concept of gene is understood functionally or defined structurally as a stretch of DNA or a nucleic acid sequence, genes are components of organized systems that can be judged to be normal or abnormal in structure or function based on their present or past contributions to evolutionary success. On Boorse's statistical "forward-looking" account, the concepts of health and disease can be defined in terms of normal and abnormal genetic functions rather than normal and abnormal physiological functions. Genome structure can be incorporated into the "species biological design" and the specific functions of genes can be identified based on their contributions to the overall functioning of individual organisms. Genes function normally if they make typical contributions to survival and reproduction among members of a particular reference class. It might be argued that by locating biological definitions for the concepts of health and disease at the level of the genome, it is no longer necessary to consider the abilities of individual organisms to survive and to reproduce. In his (1993) paper titled "Do We Need a Concept of Disease?," Germund Hesslow argues that, with increased emphasis on genic selection, the organism and its "species design" are no longer central in biology. But, as we saw in Chapter Three, the "bean bag" approach to genetics that attaches mean fitness values to single alleles in populations and ignores their interactions in specific individuals provides inadequate theoretical foundations to medicine. Clinical

interventions, at least of the sort that are associated with "genetic" or "molecular" medicine and the implementation of the new genetic technologies, are directed at individuals and not at whole populations. The relevant theoretical knowledge concerns the functions of genes and genomes in individuals. Human molecular genetics, like physiology, deals with "proximate" causes that contribute to the overall functioning of individual organisms: genes are just such causes.

On Millikan's and Neander's nonstatistical "backward-looking" functionalist accounts of health and disease, the "proper function" of a gene is identified on the basis of its evolutionary history in a particular population or species. A "normal" gene exercises the function for which it was selected in the environment in which it was selected. Disease or dysfunction arises where a gene has mutated from its "normal" form or where the "normal" gene finds itself in the "wrong" environment. These accounts are consistent with the evolutionary approach to medicine that Randolph M. Nesse and George C. Williams (1994) call "the new science of Darwinian medicine." Nesse and Williams argue that genuinely "defective" genes are infrequent causes of disease because they are maintained at low frequencies by negative selection in mutation-selection balance. Many common conditions and diseases actually reflect evolutionary adaptations confronted with a changed environment. Such conditions as vitamin deficiencies, diabetes, hypertension, myopia, alcohol and drug addiction, dental caries, and obesity are regarded by Nesse and Williams as "diseases of civilization." These result from the interactions of genes that were favoured by natural selection in our hunter-gatherer past with the "novel environments" associated with the past 10,000 years of civilization. From an evolutionary perspective, this is far too short a time for humans to have adapted to these "new" environmental circumstances. Instead, we are "specifically adapted to Stone Age conditions" (p. 134) — an "ancestral environment" that is referred to by anthropologists as "the environment of evolutionary adaptedness" or the EEA (p. 138). "Susceptibility" genes reflect adaptations that are no longer adaptive in contemporary environments. Like Millikan and Neander who regard such genes as "normal" and today's environments as "abnormal," Nesse and Williams refer to "susceptibility" genes as "quirks" of evolution rather than mutations.

Up to this point, for purposes of explication, I have emphasized the differences between "forward-looking" and "backward-looking" accounts of biological functions. The rest of the chapter, however, focuses on aspects in which they are similar. Both "forward-looking" and "backward-looking" approaches represent disease and dysfunction in terms of deviation from already-established objective functional norms. In this way, biology furnishes medicine with theoretical foundations. This not only lends medicine scientific legitimacy but serves to sanction clinical interventions that seek to restore the body (or the genome) to its "natural" condition. Each of two assumptions implicit in this approach is taken up, in turn, in the remaining two sections of the chapter.

The first assumption is that biology offers theoretical foundations to clinical medicine because knowledge of functional norms precedes knowledge of disease and dysfunction. Nineteenth-century physiologist Claude Bernard was an early proponent of the approach taken by Boorse, that is, the founding of a scientific medicine in physiology. In *Introduction to the Study of Experimental Medicine*, Bernard writes:

> Since science can be established only by the comparative method, knowledge of pathological or abnormal conditions cannot be gained without previous knowledge of normal states, just as the therapeutic action of abnormal agents, or medicines, on the organism cannot be scientifically understood without first studying the physiological action of the normal agents which maintain the phenomena of life. (1957, p. 2)

Although the "new science of Darwinian medicine" represents a late-twentieth century challenge to the hegemonic authority of physiology, it likewise assumes that the normal is epistemically prior to the pathological. In urging that psychiatry adopt the evolutionary approach of "Darwinian medicine," Nesse and Williams refer to the importance of understanding normal functions before identifying the "flaws that cause disease":

> The research findings [in psychiatry] are solid, but they are not connected in any coherent theory. In its attempt to emulate other medical research by searching for the molecular mechanisms of disease, psychiatry has ironically deprived itself of precisely the concepts that provide the tacit foundation for the rest of medical research. By trying to find the flaws that cause disease without understanding normal functions of the mechanisms, psychiatry puts the cart before the horse. (p. 230)

In fact, Nesse and Williams go on to justify this approach in psychiatry by analogy to the success of physiology's provision of theoretical foundations to the study of pathophysiology and the practice of medicine:

> an evolutionary view is psychiatry's route to genuine integration with the rest of medicine. An intensive effort to understand the functions of the emotions and how they are normally regulated would provide, for psychiatry, something comparable to what physiology provides for the rest of medicine. It would provide a framework in which pathopsychology could be studied like pathophysiology, so we can understand what has gone wrong with the normal functioning of bodily systems. (p. 232)

In section 4.2, I present Georges Canguilhem's arguments in *The Normal and the Pathological* that challenge the validity both of the general approach that founds the study and clinical treatment of disease and dysfunction in the knowledge of normal function and of the specific characterization of the relationships between physiology, pathology, and clinical medicine to which Nesse and Williams adhere.

The second assumption shared by both functionalist approaches to health and disease is that the distinctions between normal and abnormal function and health and disease are scientifically objective and escape the influence of nonbiological values, be they moral, aesthetic, social, or cultural. Boorse takes one route to objectivity and Millikan and Neander take another. Boorse's representation of "normal" functions is statistical: a particular part or process functions normally if it makes a "typical" contribution to "individual survival and reproduction ... within members of a reference group" (1977, p. 562). Functions themselves are assigned according to the "means-end hierarchies" that are discernible in "species biological designs" by empirical observation alone. In other words, Boorse believes his judgements of normal and abnormal physiological functions to be not at all normative but merely descriptions of empirical fact. In contrast, Millikan and Neander reject Boorse's statistical non-normative approach. They distinguish between what entities *actually* do and what they are *supposed* to do. Biological (functional) norms are objective and discoverable properties in nature that have been established by natural selection. Judgements concerning normal and abnormal functions — and, derivatively, health and disease — are therefore irreducibly normative. Boorse has been justly criticized for presuming that his account avoids normativity. As

Martin Bunzl (1980) points out, an account that defines the concepts of health and disease in terms of wholly empirical notions of "species design" and (statistically) normal and abnormal biological functions cannot accommodate deviations from "species design" that do not detract from, but improve upon, function. These demand the normative judgement that survival and reproduction are "goods" and that it is better for a trait to be associated with superior than with average fitness. The concepts of health and disease are rendered no less scientific or objective, however, since the norm of reproductive success is captured within the theoretical framework of evolutionary biology. George J. Agich (1983) refers to functionalist accounts of health and disease based on judgements that are both descriptive and normative as "weakly normative" where the appeal is only to norms based in scientific theory (p. 29). I adopt this term to describe both "forward-" and "backward-looking" functionalist accounts of health and disease, despite the different (epistemological versus ontological) grounds for their avowed scientific objectivity. A legitimate distinction can be drawn between normative accounts of health and disease that appeal only to biological norms and those, referred to as "strongly normative" by Agich, that incorporate nonbiological, that is, moral, aesthetic, social, or cultural, norms. In section 4.3, I argue that judgements of health and disease, normal and abnormal functions, and normal and mutant genes are always normative in this stronger sense. Judgements of normal and abnormal biological function are always relative to an environment. Insofar as human environments are irreducibly social and cultural, as well as physical and biological, so too are such judgements.

4.2 From the Pathological to the Normal and From Physiology to Human Molecular Genetics

4.2.1 Georges Canguilhem: From the Pathological to the Normal

Georges Canguilhem's thesis in *The Normal and the Pathological* is that the positivist[3]

[3] Canguilhem has in mind nineteenth-century positivists such as Auguste Comte and Claude Bernard.

adage that one "knows in order to act" is mistaken with respect to the relationship between medicine and biology. Technology is not merely the application of science. The relationship between biology and medicine is not unidirectional, consisting in medicine's practical application of the theoretical knowledge of biology. Canguilhem's thesis is two-part. He argues first that it is not the prior knowledge of biological norms that informs our theoretical understandings of disease processes and thereby directs medical intervention, but, rather, the pathological that is antecedent to and constitutive of the normal. In Canguilhem's words: "Disease reveals normal functions to us at the precise moment when it deprives us of their exercise" (1989, p. 101). Physiology is indebted to pathology, not vice versa. One identifies the normal function of a part or process when something goes wrong. In humans, diseases offer the opportunity for what Auguste Comte called "spontaneous experiments" (in Canguilhem 1989, p. 51). In experimental physiology, "artificial pathologies" are induced in laboratory animals. Canguilhem argues secondly that the science of pathology, in a similar way, arises from and is indebted to clinical medicine. Pathologists study disease in order to find suitable ways for physicians to treat patients. Clinical judgements of health and disease inform the directions research pathologists take in the laboratory. "Artificial pathology" is created in imitation of its natural counterpart. For example, although Brown-Séquard has been credited with founding endocrinology in 1856 when he caused the death of an animal by removing its adrenal gland, the experimental tack he chose is understandable only in view of Addison's 1855 description of a disease condition associated with adrenal gland attack. How laboratory data are interpreted is also influenced by clinical judgements. Statistical judgements of biological normality and abnormality, taken alone, are uninformative with respect to health and disease. Physiologists may find that 120 is a typical systolic blood pressure but physiological constants are appropriate clinical norms only to the extent that they are found to be constitutive of health. Similarly, anatomists identify structural anomalies as irregularities or statistical divergences but consider them to be pathological only if they are found to disrupt function. For example, a structural anomaly like the sacralization of L5 (the fifth lumbar vertebra) is considered innocuous unless it is associated with mechanical low back pain.

One must not interpret Canguilhem's thesis to be a historical and sociological claim about how the scientific institutions of pathology and physiology came to be established. Rather, the thesis points out the historical, logical, and material aspects of the scientific explanations of normal and abnormal function and of health and disease that are found in physiology and pathology. This is illustrated in Canguilhem's criticisms of Rudolf Virchow's ontological conception of disease, which he derisively refers to as "atomistic pathology." Virchow's belief that "the essence of disease is a modified part of the organism or a modified cell or modified aggregate of cells (or tissue or organ)" (in Canguilhem 1989, pp. 224-225) is considered by Canguilhem to involve a "selective forgetting." What is forgotten, Canguilhem says, is the very necessary role the organism, in dynamic interaction with its environment, plays with respect to the generation of scientific explanations concerning the normality or abnormality of its parts: "we forget that historically, logically, and histologically we reached the cell by moving backward, starting from the total organism; and thought, if not the gaze, was always turned toward it" (p. 223). What Canguilhem means by "histologically" is that the tissues studied in the laboratories of the physiologist or pathologist are samples taken from actual individuals who are known to be either healthy or diseased. "Historically," this information about the source of the material is incorporated into explanations concerning normal or abnormal physiology: normal cell structure and function explain the health of some individuals; abnormal cell structure and function explain the diseases of others. "Logically," this pattern of explanation that proceeds from the experience of individual organisms in their environments to knowledge of the function of their parts is necessary; it does not merely reflect what tends to happen.

In this section of the chapter, I focus on only one aspect of Canguilhem's thesis: that clinical and scientific judgements concerning disease are antecedent to and constitutive of scientific judgements concerning physiological normality. The third section of the chapter, however, attends to a second aspect of Canguilhem's thesis. The concept of "biological normativity" is primordial in Canguilhem's account. This concept represents the relationship of "dynamic polarity" that exists between individual organisms and their environments; specifically, it refers to the organism's ability to establish, on an ongoing basis, new vital norms in response to fluctuations in the environment. Health

consists in the preservation of "biological normativity" and disease is perceived by the organism as the onset of impairment in its ability to adapt to environmental challenges. Thus, judgements of health and disease are rooted in the "affective" experiences of individuals; medical judgements are always secondary. Physicians exist for patients, not vice versa: "it is first and foremost because men feel sick that a medicine exists. It is only secondarily that men know, because medicine exists, in what way they are sick" (1989, p. 229). In sum: were it not for individuals who get sick, there would be no clinical medicine; were it not for clinical medicine, there would be no study of pathology; and, were it not for the study of pathology, there would be no science of physiology. Ultimately, "[i]t is life itself and not medical judgment which makes the biological normal a concept of value and not a concept of statistical reality" (p. 131). The normativity that attaches to medical judgements of health and disease finds its source in the concrete biological experiences of individuals and not in the extent to which the beliefs of physicians have been shaped by moral, aesthetic, social, or cultural values. For Canguilhem, disease is *not* "whatever physicians in a particular society treat" (p. 33), as H. Tristram Engelhardt Jr. (1981) characterizes the social constructivist view, simply because the valuation of biological norms lies not in "the normative activity of therapeutics" (Canguilhem 1989, p. 131) but in the "dynamic polarity of life itself." It is a "facile relativism," Canguilhem contends, that denies any distinction between health and disease (p. 77).

4.2.2 *Can Canguilhem's Thesis be Extended from Physiology to Human Molecular Genetics?*

Canguilhem argues that the science of physiology "stands at the crossroads of the laboratory and the clinic" (1989, p. 111) and that insofar as it seeks to explain health and disease objectively in terms of the structures and functions of parts of organisms alone, whether these be individual organs, tissues, or cells, it forgets its debt to the clinic. Do explanations of health and disease in terms of normal and abnormal gene structure and function bear the same explanatory debt to the relationships between individuals and their environments? Is it our experiences with disease that provide access to knowledge of what is genetically normal? If Canguilhem's thesis that knowledge of the pathological

is antecedent to and constitutive of knowledge of the normal is to hold for human molecular genetics as well as for physiology, it must be established that our knowledge of normal gene structure and function depends on prior knowledge of mutant gene structure and function and that this in turn follows from clinical judgements of health and disease.

Genetics is the study of the inheritance of differences — specifically, the relationships between genetic differences and phenotypic differences. Genetic mutations have been critical to the study of heredity. In part, this is because, without genetic mutation, there would be no genetic variation and, consequently, no phenotypic variation of nonenvironmental origin. In the absence of genetic variation, genotypes are still inherited and genes are still influential in the development of organisms. Nevertheless, the modes of hereditary transmission and of gene action are not anywhere nearly so readily open to investigation. The existence of genetic variation alone, however, is insufficient for either of these to be investigated using the traditional methods of genetics. Rather, genetic variation must be associated with phenotypic variation. In the first half of this century, only genetic differences associated with differences in gross organismal phenotype were accessible to study. This dependence of the science of heredity on phenotypic differences is no doubt what the early Mendelian William Bateson had in mind when, in the collecting of plant mutants, he urged that "our exceptions be treasured" (in Cooper & Krawczak 1993, p. v). Similarly, only when T. H. Morgan made his famous 1910 discovery of a white-eyed mutant in one of his culture bottles of *Drosophila melanogaster* did classical genetics receive its start. During the period now known as classical or transmission genetics, geneticists sought to establish the basic mechanisms of hereditary transmission by tracing the patterns of inheritance of gross morphological differences, such as red eyes versus white eyes in *Drosophila*, from generation to generation. In the early days, classical geneticists had to wait for mutations to occur spontaneously in flies and these mutations could then be maintained in laboratory stocks. In 1927, H. J. Muller developed an experimental technique utilizing radiation-induced mutagenesis, and chemical mutagens were discovered not long after. Instead of these gross morphological differences studied by "the Drosophilists," biochemical geneticists relied on gross functional differences such as the ability or inability of microorganisms

like *Neurospora* to survive and to reproduce in different nutritional media in order to begin to study the basic mechanisms of gene action — the one gene-one enzyme hypothesis originated in George Beadle and Edward Tatum's work on *Neurospora* in the 1930s. Whether questions concerned hereditary transmission or gene action, genetic differences *and* differences in gross organismal phenotypes were absolutely essential in seeking to answer them.

Mid-century technological developments have in some ways changed the nature of this dependence of genetics on phenotypic differences. In the late 1940s, it became possible to detect phenotypic variation at the molecular (protein) level. Gel electrophoresis allowed some variations in the structure and function of proteins to be detected by the differences in their movements in an electrically charged field. For the first time, it became possible to distinguish heterozygote carriers of recessive alleles from non-carrier homozygote "normals." For example, in 1949, Harvey Itano, a student of Linus Pauling, discovered by electrophoresis that hemoglobin molecules taken from individuals with sickle-cell anemia carry a positive charge and that hemoglobin molecules taken from healthy individuals are negatively charged. The heterozygote status of a number of apparently healthy individuals was revealed when about half their hemoglobin was found to be positively charged and the other half negatively charged (Judson 1996, pp. 302-303). Once gross phenotypic differences can be explained in terms of differences in protein structure that are understood to be the direct consequence of differences in gene structure, it becomes possible to establish the genetic basis of a trait without relying just on its pattern of inheritance from one generation to the next. In 1956, using an enzyme to digest hemoglobin into smaller pieces, electrophoresis, and paper chromatography, Vernon Ingram found that normal and sickle-cell hemoglobin molecules differ in a single amino acid (ibid., pp. 301-307). With advances in protein-sequencing techniques, and the elucidation of the genetic code by 1967, genetic structure could be inferred from knowledge of protein structure alone. This meant that the nucleotide sequence of a gene, although not its chromosomal location, could be identified by determining the linear amino acid sequence of a single protein molecule.

Genetic mutations have been critical to the study of heredity in an additional way. Geneticists, like scientists in other fields, seek to discover fundamental laws of nature.

The Mendelian patterns of inheritance that are revealed through the study of genetic differences at single loci hold for normal (wild-type) and abnormal (mutant) alleles alike. The genetic code applies to codon nucleotide-triplets whether they comprise part of the wild-type or mutant gene. Genetic and phenotypic differences were necessary for these basic mechanisms to be discovered but none among these differences falls outside the "law," so to speak. But much genetic and phenotypic variation is not understood simply as difference but rather as deviation — deviation from what is lawful, "natural," or "normal." Since their inception, and until very recently, genetic maps have been representations of the aberrant or the pathological. Alfred H. Sturtevant constructed the first genetic map in 1913. Morgan had already discovered linkage: that white eyes and male sex are inherited together in *Drosophila*. Sturtevant, an undergraduate student at the time, had the brilliant insight that the relative distances between genes on the same chromosome could be estimated according to the frequencies with which they are inherited together rather than recombining. This would establish the degree of linkage between them and their rough proximity on the chromosome. Sturtevant's 1913 map was a horizontal line representing the X-chromosome in which vertical lines marked off the relative positions of six different genes. All of these genes represented mutant characteristics — white eyes, for example, instead of the wild-type red. The maps of "the Drosophilists" portray the relative positions of gene loci in linear arrangement on the chromosomes as revealed by the linkage between mutant alleles at these loci. Initially, classical geneticists relied on linkage between two gross organismal phenotypic traits. With the development of cytogenetic staining techniques, however, it became possible to map some phenotypic traits to regions of the chromosome where some gross aberration of the chromosome was visible under the microscope. In this way the fly's giant salivary chromosomes were a huge boon to "the Drosophilists." Since the inception of laboratory genetics, it has been mutations that have been mapped. Normal genes and chromosomes have been understood as those in which these mapped mutations are absent.

With the 1970 discovery of bacterial enzymes called restriction nucleases and their use in creating restriction fragment location polymorphisms (RFLP) maps beginning in 1980, linkage mapping could be carried out for a single variable phenotypic trait using linkage to genetic markers located in highly variable regions of the genome instead of

needing to rely on linkage to another phenotypic trait, whether organismal or molecular. Many mapping approaches in contemporary human molecular genetics remain tied, however, to linkage mapping techniques developed in classical genetics and cytogenetics. One begins with a disease phenotype, uses cytogenetic or linkage analysis to isolate and to characterize genes and their mutations, and then seeks to establish the causal pathway through messenger RNA (mRNA) to protein and back to the disease phenotype (Berg 1993). Currently used techniques that proceed from phenotype to genotype in this way include "functional cloning," the "candidate gene approach," "positional cloning," and the "positional candidate approach." "Functional cloning" identifies the gene through knowledge of the structure of a protein involved in a particular disease by applying the genetic code in reverse. Its chromosomal location is revealed secondarily by synthesizing a DNA probe that hybridizes to a genomic DNA or complementary DNA (cDNA) library. The "candidate gene approach" is similar but has only partial data available on the molecular aspects of the disease. In "positional cloning," there is no molecular data on the disease available. One begins with a group of related individuals some of whom have the disease and some of whom are healthy. Using either cytogenetic or linkage analysis, the disease trait is linked to an approximate chromosomal region. Painstaking "chromosomal walks" are then carried out to try to locate the actual gene. The "positional candidate approach" first maps the gene to a chromosomal subregion, usually by linkage analysis. Rather than trying to locate the gene directly, it accesses DNA databanks in order to survey the interval for possible candidate genes. This method became possible only as the genome map has become increasingly dense but is now overtaking the others (Collins 1995, pp. 347-348).

There are two ways in which these traditional mapping techniques used in human molecular genetics are consistent with Canguilhem's thesis that knowledge of the pathological is antecedent to and constitutive of knowledge of the normal. First, human genetics has to rely on the "spontaneous experiments" made possible by disease phenotypes because experimenters cannot selectively breed or induce mutations by radiation, chemicals, or recombinant DNA technologies in humans as they do in experimental organisms. Geneticists throughout the century have made ample use of the opportunities provided by "inbred errors of metabolism" and other hereditary diseases,

"consanguineous" (relatively "inbred") families in which there is a statistically higher frequency of rare hereditary disorders, the radiation exposure sustained by the victims of Nagasaki and Hiroshima, and the careful family records maintained by groups like the Mormons and the Amish. Second, in human and nonhuman organisms alike, only by mapping mutations associated with variant phenotypes is it possible to locate the functional segments of DNA identified as genes. Linkage mapping moves from disease phenotypes to mutant genes, and, finally, to normal genes. *Mendelian Inheritance in Man* (*MIM*) catalogues human genes — not all human genes but rather genetic "defects" or "lesions." For this reason, the author of *MIM*, human geneticist and physician Victor A. McKusick, describes the catalogues as "the morbid anatomy of the human genome" and a "diagnostic biopsy of the human genome" (1992, p. xxix). Gene maps, until very recently, have been "mutant" or "morbid" maps. Normal genes are functional segments of DNA in which mutations are absent.

This traditional genetic paradigm that identifies normal genes by first locating mutations associated with various disease phenotypes has been supplanted to a great extent by the techniques of "reverse genetics." "Reverse genetics" proceeds from genotype to phenotype rather than from phenotype to genotype:

> the reverse genetics paradigm begins with the gene as a segment of DNA whose molecular structure is known and proceeds to explore the gene's contribution to the organism's phenotype; thus, the experimental path is from the gene as a nucleotide sequence to the corresponding phenotypic characteristic. (Berg 1993, p. 263)

In this reversal, the new techniques of "reverse genetics" appear to challenge the extension of Canguilhem's thesis to human molecular genetics. Because of the clues provided by nucleotide sequences commonly found in coding regions, it is possible to locate genes by rapid ("blind") sequencing of genomic DNA. Genes can also be mapped and characterized by using polymerase chain reaction (PCR) to amplify mRNAs taken from various tissues, obtaining complementary DNA (cDNA) through reverse transciption, and mapping cDNAs to chromosomal locations by using DNA probes and *in situ* hybridization. This means that "normal" or "wild-type" genes can be mapped and sequenced directly without needing first to locate their mutant variants. Even genes that do not vary within a species can be mapped and sequenced. Provided that the DNA or

mRNA derives from samples taken from healthy individuals, one finds "normal" genes. Similarly, functional DNA sequences found in "normal" or "wild-type" experimental organisms may be considered normal for that species and, perhaps, if they serve a basic cellular function, many species.

But to determine what the function of a particular — presumably normal — segment of DNA actually is, clinical studies or laboratory experiments are essential. Clinical data can be used to identify normal and mutant gene structures by comparing DNA or mRNA sequences obtained from normal and cancerous tissues or from control subjects and patients with diseases. Functional knowledge is achieved when the specific ways in which things have gone wrong in diseased individuals help to reveal normal functions. For "reverse" as well as "forward genetics," knowledge of abnormal functions precedes knowledge of normal functions and genes are identified as normal or abnormal only as the result of antecedent clinical judgements of health and disease at the level of individual phenotypes. But the "reverse genetics" approach, with its recombinant DNA tool-kit, is no longer dependent on the spontaneous mutations associated with human disease phenotypes or limited to undirected radiation- or chemical-induced mutations in laboratory genetics. Recombinant DNA technologies make it possible to use "highly directed" modifications of DNA to attempt to discover gene function. Nucleotide bases can be "knocked out" or inserted into genes and normal functions determined by observing phenotypic changes at various levels: "in vitro, in cultured cells, or ... in whole organisms" (Berg 1993, p. 263). In the laboratory, as in the clinic, knowledge of the pathological is antecedent to knowledge of the normal. Gene functions are only uncovered in the event that noticeable phenotypic changes result from the experimental production of mutations. Where a stretch of DNA or a single nucleotide base is modified or "knocked-out," in the absence of discernible phenotypic changes, the DNA is assumed to be either functionless, functionally redundant, or of little functional importance.

Although it is necessary to create "artificial pathologies" in the laboratory in order to investigate normal gene functions, the explanations of normal and pathological functions that are generated seem to bear no historical dependence on prior clinical judgements of health and disease at the level of individual phenotypes. Thus, while it seems possible to extend the first part of Canguilhem's thesis — that knowledge of the

pathological is antecedent to and constitutive of knowledge of the normal — from physiology to human molecular genetics, the extension of the second part of the thesis — that knowledge of the pathological is directed by prior clinical judgements of health and disease — seems questionable. However, as Canguilhem remarks about Bernard's experimental physiology, laboratory norms can be considered to apply to humans and to be relevant to the treatment of disease only to the extent to which experimental organisms and humans, and experimentally induced and spontaneously occurring pathological states, are similar. That human molecular genetics maintains this aim to create "artificial pathology" in imitation of its natural counterpart is reflected in the current use of various experimental organisms as "model systems" for the study of human diseases. Due to common evolutionary origins, homologous genes in many "simple" species are helpful to uncovering basic cellular functions common to all. Cloned human genes can also be inserted into the germ-lines of experimental organisms bred for research: transgenic mutant strains of mice are popular models for human diseases.[4] Taken alone, "objective" scientific norms do not dictate directions for clinical practice; rather, such norms are accepted as guides to practical interventions according to their correspondence with antecedent clinical judgements of health and disease. Despite initial appearances to the contrary, the second part of Canguilhem's thesis — that knowledge of the pathological is directed by prior clinical judgements of health and disease — appears to be true of "reverse genetics" as well as "forward genetics."

In this way, what Canguilhem has written about physiology seems true for human molecular genetics: for "forward" and "reverse" genetics alike, "historically, logically, and histologically we reached the [gene] by moving backward, starting from the total organism; and thought, if not the gaze, was always turned toward it" (p. 223). "Histologically," normal gene structure and function are identified by using DNA or mRNA samples known to originate in healthy or diseased individuals or, in the laboratory, by using mutant DNA or RNA sequences that have been obtained by modifying "wild-type" or normal sequences. "Historically," knowledge of material's source becomes incorporated into explanations concerning normal and abnormal gene structure and

[4] For more on transgenic animals, see Jaenisch (1988) and Rusconi (1996).

function. In the case of clinical data, one distinguishes between normal and abnormal genes on the basis of a prior distinction between normal and abnormal phenotypes. In the laboratory, "thought is always turned toward the complete organism" and clinical states of health and disease insofar as biological knowledge is to be applied to clinical practice. "Logically," do clinical judgements and knowledge of the pathological *tend* to precede knowledge of normal function or is the pathological *necessarily* antecedent to and constitutive of the normal? Might the heading in a cell biology textbook that informs students that "Mutant Organisms Best Reveal the Function of a Gene" (Alberts et al. 1998, p. 339) be rewritten as "Only Mutant Organisms Reveal the Function of a Gene"?

Even using the traditional techniques of "forward genetics," why can an objective standard of genetic normality not be established by mapping normal genes? Where there is no phenotypic variation in a trait, genes that influence the trait's expression cannot be mapped except by experimental manipulation. It is therefore impossible, using the techniques of "forward genetics," to map either traits that are universal to a species or reference class or functionally equivalent "normal" gene variants with indistinguishable phenotypic effects. Variation in some human phenotypic traits — for example, eye colour, hair colour, or, within limits, height — is considered to be normal. Brown eyes and blue eyes, after all, see equally well. If green eyes are considered to be abnormal, this is only in the statistical sense that they are rare. Normal phenotypic variation is either qualitative/discontinuous or quantitative/continuous. Qualitative/discontinuous variation is associated with differences in single genes. Quantitative/continuous variation is associated with differences in multiple genetic and/or nongenetic factors. Insofar as it is possible to map "normal" gene variants that are associated with "normal" variation in qualitative and/or quantitative phenotypic traits, it seems that we would be forced to conclude that knowledge of the normal is not *necessarily* dependent on either prior knowledge of the pathological or antecedent clinical judgements of health and disease.

It *is* possible to map "normal" gene variants associated with "normal" variation in qualitative phenotypic traits using the traditional approach of "forward genetics." Such "normal" qualitative/discontinuous traits as PTC-tasting and eye colour are often referred to as "neutral" or "normal" polymorphisms. It is not because these variants are

functionally indifferent that they are considered to be "neutral" or "normal." There are physiological differences: one is able to, or not able to, taste PTC; one is able to, or not able to, produce melanin in large quantities. The inability to taste PTC or to produce melanin may even be considered to be functional losses if these abilities are known to have prevailed in ancestral populations. The only reason that genes associated with the loss of the ability to taste PTC or to produce melanin are considered to be "normal" variants and genes associated with the loss of the ability to hear are considered to be "abnormal" mutants is that these physiological functions have different meanings at the level of the organism. What counts as normal or abnormal genetic variation, difference or deviation, depends ultimately on what is judged to be normal or abnormal, healthy or diseased, in actual individuals in particular environments. When we call a gene that lacks a mutation associated with congenital deafness "normal," there is an antecedent judgement that the inability to hear, unlike the inability to taste PTC or to produce melanin, represents an abnormal or pathological functional loss. Hence, the mapping of "normal" gene variants associated with qualitative phenotypic differences follows from clinical judgements of health and disease and is consistent with the second part of Canguilhem's thesis. As for the first part of Canguilhem's thesis, that knowledge of the pathological is antecedent to and constitutive of the normal, it needs to be emphasized that what has been identified is a *normal* gene *variant* and not a *normal* gene *function*. We might say that the gene is *involved* in PTC-tasting: this ability is permitted by one allele and precluded by another. We would not say that the gene's normal function, or even its function, *is* PTC-tasting. To do so, the inability of organisms to taste PTC would have to be in some way maladaptive which, at least today, it is not. Knowledge of the pathological is indeed antecedent to knowledge of normal function.

Apart from these polymorphisms, the traditional techniques of "forward genetics" have been limited to the mapping of "mutant" traits associated with single gene differences because most "normal" traits have multiple genetic and nongenetic determinants. Although the presence of a single gene may be correlated highly with the presence of a trait such as congenital deafness, the ability to hear supports no such correlations. Consequently, correlation-dependent linkage mapping is impossible. Only by locating "mutant" genes associated with deafness is it possible to identify "normal"

genes involved in hearing.[5] Two recent technological developments challenge the limitations of classical monogenic mapping.

First, the availability of dense genetic maps in a variety of species including humans has resulted in the ability to map quantitative trait loci (QTLs). Mapping is no longer restricted to traits associated with large single gene "effects." QTLs are loci that are believed to have varying degrees of influence on complex traits, although actual genes have not yet been cloned (Paterson 1998). QTLs are identified by finding associations in large experimental populations between variation at genetic marker loci and phenotypic variation in the trait of interest. QTLs are labelled as "+" or "-" depending on whether they increase or decrease the value of the trait in question and this expression will presumably be modulated by different alleles at the loci. Complex traits being mapped in humans include "susceptibilities" to conditions such as cancer, cardiovascular disease, obesity, schizophrenia, diabetes, and asthma. "Normal" traits that vary continuously among individuals may also yield to "molecular dissection": personality, intelligence, size, etc. It is possible, as we saw in Chapter Two, to consider the entire normal distribution curve of values for a continuously varying trait in a population to be the adaptive norm. Any individual who falls within the distribution's range of values is considered to be "normal." However, one or both tails of the distributions of many continuously varying traits are often associated with pathology — for example, body height, blood hormone levels, or brain neurotransmitter concentrations. Whether the entire distribution of values is considered to be normal and, if it is not, where the line is drawn between normal and abnormal variation depend on antecedent clinical judgements of health and disease. More or less matters when it comes to "normal" variation in traits such as shyness and intelligence. As with the PTC-tasting example, if all phenotypes associated with variation at a particular QTL are considered to be normal or healthy, while we might discover the genetic mechanisms that underlie identifiable physiological differences, we cannot identify normal gene functions unless there exists genetic and

[5] This is a point that is frequently made by critics of genetic determinism: "one can break a transistor radio by removing one component, but no one would seriously argue that the missing component alone normally causes the radio to play a particular radio station" (Berkowitz 1996, p. 46).

phenotypic variation that is considered to be pathological. Let us assume that there is a regulatory gene that contributes to quantitative variation in levels of aggression found among members of a population. This regulatory gene is associated with the transcription of an enzyme that catalyzes a hormone linked to aggressive behaviour into its less potent substrates. The enzyme's rate of transcription depends on the particular allele that is present at the regulatory gene locus as well as cellular concentrations of the hormone. In some individuals, but not in others, increased hormonal concentrations are associated with elevated transcription rates and the degree of this effect is correlated with levels of aggression exhibited. However, there is no basis for concluding that the transcription effect is the gene's normal function if all individuals are considered to be normal and healthy, even those in whom the effect is completely absent. Knowledge of the pathological is necessarily antecedent to knowledge of normal function.

The second technological development that challenges the limitations of classical monogenic mapping is "reverse genetics." Using the techniques of "reverse genetics," "normal" or "wild-type" genes, even genes that are universal to a species or reference class, can be mapped. Although, as we have already seen, the functions of mapped "wild-type" genes are revealed only through spontaneous or induced pathological states, it is possible to map and to sequence normal genes by using DNA or mRNA that has been obtained from normal healthy individuals or "wild-type" experimental organisms. There are problems with this, however. "Wild-type" laboratory strains of organisms are quite atypical in that they are highly inbred and therefore genetically very homogeneous. In attempts to infer "consensus sequences" for given functions in the laboratory, it must be remembered that, in "nature," considerable genetic variation at the same and at different loci may be associated with performance of a function. Similarly, it cannot be assumed that a single nucleotide sequence obtained by analyzing DNA or mRNA taken from a small number of healthy individuals represents the only possible "normal" sequence. A number of functionally equivalent variants often occur at the same locus. Population data are helpful in uncovering this variability but, without accompanying functional information on each member of the population, it is impossible to distinguish an allele that is statistically rare as the result of negative selection from one that is fully functional but rare due to stochastic fluctuations of selectively neutral alleles. Ultimately,

judgements of health and disease at the level of phenotype are necessary to delimit the range of "normal" genetic variation. And, as I have already argued, it is possible to gain knowledge of normal gene functions using the techniques of "reverse genetics" only with the help of clinical and/or experimental studies which offer antecedent knowledge of actual or "artificial" pathologies.

In contemporary human molecular genetics, knowledge of normal and abnormal gene structure and function is obtained by combining the techniques of "forward" and "reverse" genetics. Oscillation occurs between the levels of genotype and phenotype. Scientific knowledge of normal gene function does not, however, precede scientific knowledge of abnormal gene function. Rather, the pathological is antecedent to and constitutive of the normal. Distinctions between normal and abnormal genes and genotypes are themselves indebted to antecedent clinical judgements concerning what counts as health and disease at the level of individual phenotypes. This casts doubt on the ability of human molecular genetics to provide value-neutral theoretical foundations for diagnoses of health and disease in "genetic medicine" and for the propriety of clinical interventions using genetic technologies that aim to restore what is "normal" or "natural." But it in no way entails that the distinctions between normal and abnormal genetic mutation and genetic normality and genetic mutation are idly invented or are "strongly" normative, that is, governed by moral, aesthetic, social, or cultural values. The distinctions between normal and abnormal genetic variation and genetic normality and genetic mutation are neither more nor less value-laden than the clinical judgements of health and disease that precede and determine them. What counts as a disease matters because this determines what counts as a mutation and ultimately, therefore, what counts as normal. The normative status of the clinical judgements that distinguish between health and disease is the topic of the final section of the chapter.

4.3 Genetic Normality and Genetic Mutation: Negotiating Their Social and Cultural Meanings

I have argued that Canguilhem's two-part thesis can be extended from physiology to human molecular genetics. First, it is not the prior knowledge of biological norms that informs our theoretical understandings of disease processes and thereby directs medical intervention, but, rather, the pathological which is antecedent to and constitutive of the normal. Second, the science of pathology, in a similar way, arises from and is indebted to clinical medicine. In other words, clinical judgements of health and disease precede and determine scientific judgements of the pathological and the normal. This means that any values that attach to clinical judgements of health and disease at the level of phenotype are imported to the level of the genome. Without the interpretations that are lent them from the phenotypic level, DNA sequences are meaningless chains of cytosine, guanine, thymine, and adenine bases. Unravelling the DNA "code" provides no radical new insights into what it is to be a normal or abnormal, healthy or diseased human being; rather, its meaning will merely reflect whatever are our current understandings of health and disease. Recall the second assumption that underlies the functionalist accounts of health and disease covered in the first section of this chapter: judgements of health and disease are scientifically objective and not influenced by moral, aesthetic, social, or cultural values. In this final section of the chapter, I argue that we do not and cannot understand health and disease in entirely, or even in primarily, biological terms. In criticism of Christopher Boorse's functionalist account of health and disease, George J. Agich (1983) writes: "The language of disease necessarily involves evaluation and value judgement about what comprises the proper or desirable human condition" (pp. 37-38). Moral, aesthetic, social, and cultural values inform judgements about "what comprises the proper or desirable human condition" and what counts as health or disease. Consequently, the distinctions between normal and abnormal genetic variation and genetic normality and genetic mutation are similarly value-laden.

All of Boorse, Millikan, Neander, and Canguilhem would disagree with this conclusion. Boorse contends that health and disease are nonnormative descriptive concepts that have only empirical content. Recall that, for Boorse, a healthy individual

has a capacity to survive and to reproduce that is typical among members of a reference class. On this view, the concepts of genetic normality and genetic mutation are similarly nonnormative, descriptive, and empirical. Millikan, Neander, and Canguilhem are in agreement that Boorse's statistical approach is mistaken. They believe that biological norms are implicated in determinations of health and disease. This means that the concepts of genetic normality and genetic mutation are "weakly" normative insofar as they appeal to biological — but not moral, aesthetic, social, or cultural — norms. Boorse's assumption of nonnormativity is also mistaken because, as I argued earlier, his "forward-looking" functionalist account of health and disease cannot accommodate modifications to "species design" that improve upon function unless it accepts that survival and reproduction are norms (Bunzl 1980). Although these theorists alike argue that clinical judgements of health and disease are based in biological norms, distinctly different approaches are taken by Boorse, Millikan, and Neander, on the one side, and Canguilhem, on the other. I will explain these differences in terms Canguilhem sets out in his contention that the categories of health and disease are "biologically technical and subjective, not biologically scientific and objective" (p. 222).

The accounts of Boorse, Millikan, and Neander are "biologically scientific and objective" because they are grounded in the scientific theory of evolution by natural selection. Although the biological norms appealed to — survival and reproduction — are instantiated in individuals, they are populational. The extent to which given traits contribute to the abilities of organisms to survive and to reproduce in present environments, or contributed to these abilities in past environments, is measurable only relative to other members of the population or species that lack the trait. Fitness — or relative reproductive success — is a property of individuals only as members of a population. On these accounts, clinical judgements of health and disease can be considered to be scientifically objective, as I have been using the terms 'scientifically objective,' insofar as they are informed only by biological norms (survival and reproduction) and not by moral, aesthetic, social, or cultural norms. For Boorse, genetic diseases can be characterized in terms of statistically lower survival and reproductive rates in individuals with a certain genotype. For Millikan and Neander, genetic diseases are present where "defects" or mutations in adapted genes arise or where the ancestral

environment has changed (according to Nesse and Williams, the "environment of evolutionary adaptation" or "EEA" is the "African savannah" in Stone Age times). Canguilhem means something slightly different by "biologically scientific and objective," but these senses apply as well to the functionalist accounts of health and disease found in Boorse, Millikan, and Neander. For Canguilhem, "biologically scientific" refers to the theoretical grounds that clinical judgements of health and disease find in biological theories of functions. "Objective" refers to basing clinical judgements of health and disease on individual differences within a population or species.

Canguilhem agrees that clinical judgements of health and disease are based in biological norms. He argues, however, that the source of this normativity lies not in scientific or medical judgements but ultimately in a "biological normativity" that attaches to the concrete experiences of individuals in their environments. "Biological normativity" is the dynamic ability of organisms to establish new norms in response to challenges posed by their environments. It is this primordial concept that renders the categories of health and disease "biologically technical and subjective." "Subjective" does not mean that individuals have the capacities freely to decide norms: these are determined by the evolutionary history of the population or species. It means, rather, that health and disease are properties of individuals in their immediate environments and not of individuals only as members of populations in competition to survive and to reproduce. It means also that health and disease are states that individuals *experience* and not properties they are judged by others to have. "Health is a margin of tolerance for the inconstancies of the environment," writes Canguilhem (1989, p. 197). Disease arises when this "margin of tolerance" is impaired with consequent feelings of discomfort, pain, functional loss, and impotence. "Biologically technical" refers to how perceptions of health and disease arise out of humans using their bodies to further their technological desires to dominate their environments. For example, myopia will be considered an abnormal or undesirable condition by someone who wishes to learn to fly, but not by everyone. "Perfect health" is an ideal: an "assurance in life to which no limit is fixed" (p. 201).[6] "Biological

[6] This positive conception of health coincides with the holistic definition of health adopted by the World Health Organization (1946) and is shared by theorists such as Nordenfelt (1993), Pörn (1993), and Whitbeck (1981). Health, on the one hand, is the capacity that supports one's goals

normativity," as the "dynamic polarity of life itself," is what anchors the individual perceptions of health and disease that direct the medical judgements of health and disease from which theoretical knowledge of the pathological and, finally, the normal arises.

I believe that all of these accounts of health and disease — found in Boorse, Millikan, Neander, and Canguilhem alike — inadequately attend to the social and cultural contexts in which judgements of health and disease, whether these are "objective" (population-based) or "subjective" (individual-based), are made. The appeals to the biological norms of survival and reproduction found in Boorse, Millikan, and Neander are mistaken in two ways. First, reproductive success in human populations cannot be understood in entirely biological terms because present and ancestral human environments alike are social and cultural as well as physical and biological. All organisms interact with their environments and in doing so create new environments. But the capacity of humans to modify their environments is huge. Humans make culturally specific technological choices — for example, innovations such as tools, clothing, housing, farming, industry, sanitation, genetic engineering, etc. — that affect relative reproductive success and the frequencies of alleles that are passed on to subsequent generations. Human environments also include culturally-specific moral, aesthetic, and social values that influence reproductive success through mate choice. Many traits disvalued by society, treated by clinicians, and researched by molecular geneticists involve structural differences that affect physical appearance but do not directly impair biological capacities to survive and to reproduce. However, given culturally based aesthetic preferences for some types of appearances over others, such individuals may reproduce less successfully on average than those whose appearances are held in greater aesthetic regard. Second, the legitimacy of adopting the biological norm of reproductive success, although not survival, as a medical norm must be questioned. To accept this is tantamount to saying that humans *ought* to reproduce. But humans are social and cultural, as well as biological, beings. Humans have aims, interests, and aspirations that have little to do with reproduction; many of us, in fact, take great care not to reproduce. Those who seek

---

and aspirations (Whitbeck), a sense of "well-being" (Nordenfelt), or a state of "general adaptedness" (Pörn); diseases, on the other hand, are bodily or mental processes that detract from the healthy state.

medical attention for infertility desire to parent biological offspring for diverse social reasons that are shaped by moral and cultural values. What infertility means to given individuals depends on the cultures or subcultures to which they belong and the range of alternative pursuits that are, and that they perceive to be, open to them. The significance of infertility is understood differently by those affected, as well as others, depending on the social categories to which they belong — categories like gender, class, race, ethnicity, or religion. In short, we cannot talk about reproductive success in human populations without recourse to moral, aesthetic, social, and cultural values.

Canguilhem, on the other hand, explicitly recognizes the effects of culture on biological norms. Insofar as humans construct their environments through technological choices, culture influences the course of evolution. In his later writings,[7] Canguilhem came increasingly to emphasize that human environments are social as well as physical and that "biological normativity" is influenced by the social. Psychosomatic illnesses and the reactions to stress studied by Hans Selye attest to this. Perhaps, Canguilhem writes, physiology is more an applied than a pure science, "the biological study of man in cultural situations" rather than the "the science of the functions of normal man.... [as] the man of nature" (1989, p. 271). Notwithstanding this attention to society and culture, Canguilhem's account expresses a universality based in his characterization of "biological normativity" as the property of "life itself." Health and disease are "subjective" because the property of life belongs to individual organisms and not because they arise out of the culturally-mediated phenomenological experiences of individual subjects. This is the basis for a comparison Michel Foucault makes in his introduction to Canguilhem's *The Normal and the Pathological* where he contrasts Canguilhem's "vital rationalism" — "a philosophy of knowledge, of rationality and of concept" — with its opposing strain in twentieth-century French thought — the philosophy "of experience, of sense and of subject" that is associated with Sartre and Merleau-Ponty (1989, p. 8).

---

[7] Canguilhem (1989) includes his (1943) doctoral dissertation, *Essay on the Normal and the Pathological*, and three additional essays, one of which, written in 1963, reflects back on the dissertation.

Without the philosophy "of experience, of sense and of subject," I believe that it is impossible to appreciate the extents to which social and cultural norms structure the ways in which we *interpret* even our biological experiences and to which such interpretations influence our perceptions of whether we are healthy or diseased. For Canguilhem, the distinction between health and disease rests ultimately in the concrete experiences of individuals in their environments. Although he recognizes these environments to be social as well as physical, he considers the concrete experiences that lead us to consider ourselves to be healthy or diseased to be wholly biological. There is no recognition that individuals interpret their biological experiences at the level of consciousness and that, insofar as these interpretations are linguistic, they are mediated by social and cultural norms. Individuals "call" on the doctor not only because "biological normativity" — their relationship of "dynamic polarity" with the environment — has been breached but because they interpret their condition to be unfavourable. The importance of social and cultural norms in individual perceptions of health and disease is evident in the fuzziness of the line that separates structural anomalies from disease.

Recall that, for Canguilhem, structural anomalies are considered pathological only if they impair "biological normativity." Since "biological normativity" is a property of individuals in their environments, the individual is judge of the point at which anomaly shades into disease: "An anomaly manifests itself in spatial multiplicity, disease, in chronological succession" (1989, p. 138). Consider a congenital malformation — hands without fourth and fifth digits, perhaps. Acccording to Canguilhem, since there is no pain, the condition is anomalous but not pathological because there has been no breach of "biological normativity" in the individual in whom it has always existed. In someone who lost the same two fingers due to an accident at work, the condition is pathological and not anomalous because in this case the person is aware of the limitations the injury poses. Canguilhem admits some grey area. Persons with congenital malformations may eventually perceive their relative inabilities as they compare themselves to others. In any case, any disvalue that attaches to structural anomalies — whether its sources are individuals with congenital malformations who recognize their loss relative to others or the projections of those who have lost the function of a limb through injury or accident or those who imagine the impact of such a loss — is imported into scientific explanations

of the anomalies. The embryologist, in discovering by experiment the cause of an anomaly and the path by which it arises, "convert[s]," Canguilhem writes, "anomaly into disease" (p. 139). What an anomaly means to a given individual can only be understood, however, in relation to the preexisting social and cultural expectations that Canguilhem's account ignores. Extensive statistical variation in traits underlies human diversity: people come in a variety of sizes and shapes and are more or less musical, compassionate, artistic, mathematical, energetic, happy, athletic, expressive, absent-minded, etc. Relative to others, everyone faces limitations. We judge ourselves as well as others according to cultural valuations of this variability and these limitations. As molecular techniques increasingly permit the identifications of genes associated with "normal" continuous variation, what is to stop a larger than average nose, for example, from becoming, like the congenital malformation, a "defect" rather than a statistical anomaly?

Philosophers who have abandoned functionalist accounts of health and disease have gone in several different directions. One direction has been to disclaim any need for a philosophical account of health and disease. This is the approach of many bioethicists who accept the designations of health and disease arrived at by clinicians and are concerned only with the ethical warrant of specific technological interventions and clinical practices. Philosopher of science Germund Hesslow (1993) applauds this trend in a paper in which he argues that the applied approach of bioethics is likely to prove more germane in the resolution of current ethical dilemmas than the traditional theoretical approaches of philosophers of medicine. Philip Kitcher (1996), also a philosopher of science, agrees with Hesslow that debates between objectivists and social constructivists over the concepts of health and disease are unhelpful in informing the "eugenic decisions" that the implementation of the new genetic technologies demands. On one side, objective definitions of health and disease in terms of normal and abnormal biological functions fail to appreciate nonreproductive human aims and aspirations. On the other side, Kitcher finds "social constructivism" entirely unpalatable because he believes that its "relativism" leaves us unable to recognize the difference between "eradicating Tay-Sachs in North America and extirpating the females of Northern India" (p. 216). Although I agree with him about functionalist accounts of health and disease, I believe that he is wrong about social constructivism; I will return to this point later in the chapter. Where Hesslow

rejects theory, however, Kitcher recommends that a theoretical account of "quality of life" replace the traditional attempts in philosophy of medicine to find theoretical definitions for the concepts of health and disease.

Kitcher articulates a "theoretical eugenics" that is based in a normative, but nevertheless objective, formulation of "quality of life." His account leaves parents free to choose the genetic characteristics of their offspring but urges that this freedom be exercised by engaging in "responsible procreation," a rational decision-making process that considers the qualities of lives of not only the prospective individual but others implicated in the decision — whether they be members of the family or, since funds for social programs are limited, of society at large. Kitcher emphasizes that although such judgements are evaluative, they are still objective. Judging quality of life proceeds along three dimensions. The first dimension assesses an individual's (or prospective individual's) ability to form a conception of what matters in her or his life. This involves considering the development (or probable development) of a conception of self, the (probable) maturity of this conception, and the (probable) range of possible conceptions of self available. The second dimension assesses the extent to which central desires are (or are likely to be) satisfied. For example, in prenatal screening, one concludes that those born with Tay-Sachs disease will be unable to develop a sense of self; that those born with trisomy-21 (Down's syndrome) will not develop a mature sense of self; and that those born with myotonic dystrophy, infertility, or genital malformation will have limited life possibilities with central desires likely to be thwarted. Judgements along these first two dimensions are objective; they are separable from and trump judgements along the third dimension, the subjective experience of pleasure and pain. Thus, neither the presence of pleasure nor the absence of pain is either necessary or sufficient for a life of acceptable quality.

Kitcher considers the judgements of "quality of life" upon which his "theoretical eugenics" is based to be "objective" in two senses. First, judgements of "quality of life" are universal and not relative to particular societies or cultures. According to Kitcher, "social values are only pertinent to the extent that they reflect determinants of the quality of lives" (1996, p. 160). If a judgement regarding quality of life "indicts the social milieu, not the genotype or the trait" (p. 161), we ought instead to seek to change "the

social and environmental conditions that artificially cramp the quality of lives that might have blossomed" (p. 161). Kitcher stresses the importance of ensuring that judgements of "quality of life" do not become "the arrogant judgements of an elite group" (p. 192) that are subject to "background social prejudices" such as the "[e]litist differentiations ... that favor those who are athletic, intelligent, good-looking, and well-adjusted" (p. 235). Second, judgements of "quality of life" are "objective" because they do not depend on the subjective experiences of affected individuals, for example, experiences of pain or pleasure. Kitcher believes that one can objectively assess the presence of disability or the acceptability of quality of life in the absence of input from the individual whose life or disability it is. Kitcher's example here is the "devoted family man" who believes his central desires to be fulfilled. Yet, unbeknownst to him, he is "widely regarded as a shallow, sentimental buffoon whose wife is routinely unfaithful and whose children are indifferent to him" (p. 294). While the man subjectively believes his life to be of good quality, objectively, we know otherwise.

Let us look at some conditions that present unsatisfactory "qualities of lives" along Kitcher's second dimension: the likelihood that central desires will be satisfied. Two of Kitcher's examples of cases in which central desires are likely to be thwarted are infertility and genital malformation. Consider two pseudo-hermaphroditic conditions.[8] Androgen insensitivity syndrome (AIS) involves a mutation in an X-linked gene that codes for a receptor for testosterone and dihydrotestosterone. As a result of the failure of target organs to respond to stimulation by testicular hormones, the external appearance of an individual with AIS, a genetic (XY) male, is female. Internally, testes are present, there is no uterus, and the vagina is shallow and blind-ending. Infertility results. Congenital adrenocortical hyperplasia (CAH) involves the over-production of androgens by the fetal adrenal gland due to a decrease or absence of one of the enzymes involved in the synthesis of cortisol by the adrenal cortex. Excessive levels of androgens during the process of sex differentiation in the genetically female (XX) fetus result in the

[8] My descriptions of the pseudo-hermaphroditic conditions presented in this section rely on volumes edited by Josso (1981), Forest (1989), and James (1992).

virilization of the urogenital sinus and the external genitalia, while development of the uterus and fallopian tubes are normal. Genital malformation results.

I believe that these examples illustrate the inadequacies of Kitcher's account in recognizing the extents to which moral, aesthetic, social, and cultural values influence judgements not only of health and disease but of "quality of life." Kitcher holds that judgements of "quality of life" may "indict" either the "social milieu" or the genotype or trait and that this directs the appropriate site of intervention. Is it really the case, though, that the "social milieu" is blameless for the "limited life possibilities" faced by individuals with AIS or CAH — life possibilities conceived to be so limited that it would be better not to exist at all? Kitcher's own language is instructive on this point. Specifically, he refers to conditions involving infertility as those that "make it impossible for *women* to bear children" and conditions involving genital malformation as those that "preclude *normal* sexual relations" (1996, p. 289; my italics). Presumably, since he has rejected the validity of functionalist accounts of health and disease based in reproductive fitness, it is the social values of motherhood and penile-vaginal sex that are Kitcher's concern. Women who do not have biological children and heterosexuals, homosexuals, and bisexuals who choose other forms of sexual expression may resist the claim that their lives are impoverished. It is interesting that Kitcher's example singles out infertile *women* rather than men. It seems that he has himself incorporated the "background social prejudice" that it is more important for woman than for men to have children, perhaps because it is assumed that there are other valuable ways for men to contribute to society and to achieve personal fulfilment. Ignored as well by Kitcher are arguments by theorists such as Michel Foucault (1990) and Judith Butler (1990) that scientific/medical beliefs in sex binarism are produced by societal power/knowledge structures that function to reinforce and to perpetuate heterosexual and reproductive norms.

I do agree with Kitcher on several important points. It is certainly the case that practical decisions about how the new genetic technologies are to be implemented must explicitly be recognized to be normative and not "necessitated" by the need to restore what is "natural" as in the functionalist accounts of health and disease. Biological "fact" cannot be the authority that determines the worth of individual human lives. I concur wholeheartedly also with the emphases his "utopian" framework for eugenic decision-

making places on the "widespread public discussion of values and of the social consequences of individual decisions" and a "universally-shared respect for difference" (1996, p. 202). I am sceptical, however, that these very worthy aims can be optimally facilitated by continuing to worship at the philosophical altar of "objectivity." The assumption that judgements of "quality of life" are universal ignores, as do functionalist accounts of health and disease, that *all* traits and genotypes have meanings only in social contexts and that these contexts vary from culture to culture. What counts as health and disease or acceptable and unacceptable qualities of lives, and, by extension, genetic normality and genetic mutation or acceptable and unacceptable limits of genetic variation, is not objective discoverable fact but a matter of cultural negotiation. And what is negotiated is negotiable, open to discussion, and amenable to change. It is a serious mistake to exclude from "discussion" the voices of those who are most intimately affected by genetic "diseases" out of a preference for an "objective" dispassionate rationality that is dismissive of "subjective" knowledge.

Any suitable account of health and disease or "quality of life" must recognize that these are judgements of value that incorporate moral, aesthetic, social, and cultural, as well as biological, norms. Social constructivist accounts of health and disease are an appropriate point at which to begin. A well-known "social constructivist" is H. Tristram Engelhardt, Jr. For Engelhardt, the concept of disease is always practical as well as theoretical, evaluative as well as explanatory. Because a disease designation indicates "a state of affairs as undesirable and to be overcome" (1981, p. 33), it not only names but also "enjoins to action." Since what may be considered an "undesirable state of affairs" depends on social and cultural factors as well as biological ones, disease categories are relative to particular societies and cultures:

> The concept of disease is a general scheme for explaining, predicting, and controlling dimensions of the human condition. It grades into other concepts which are political, social, educational, and moral.... Disease [is] whatever physicians in a particular society treat. (pp. 32-33)

One of Engelhardt's examples makes particularly evident how medical judgements can be influenced by political purposes and moral values: the mid-19th century disease of drapetomania (the running away of slaves). Perhaps because of such extreme examples, the wrong morals have been drawn and lessons taken from social constructivist accounts

of disease. It is a naive form of social constructivism indeed that considers diseases to be socially invented fictions without "real" physically instantiated signs and symptoms. To say that all biological experiences, functions, and "facts" are culturally interpreted as well as, in humans, the products of both biology and culture is in no way to efface the biological. Meaning is negotiated in particular social contexts. Biology and culture are interdependent and co-constitutive. They ought not be placed into an opposition that assumes that what is biological cannot be cultural and vice versa.

This interdependence can be appreciated from the directions both of biology and of culture. A basic biological function like menstruation receives cultural interpretation. To say that different cultures attach different meanings to menstruation — it is a secret to be hidden; it is a mark of power; it is to be celebrated; it is an evil curse — is not to deny the very real physical occurrence of monthly bleeding. From the direction of culture, it seems that moving to the level of the genome to explain health and disease would make it far easier to reveal a diagnosis like drapetomania to be a fictional entity invented by powerful members of a slave-owning society to serve entirely political and economic purposes. Finding a "genetic basis" for a particular disease would seem to establish its legitimacy and physical reality, whereas repeated failures to locate a "disease" gene might erode confidence in the organic basis of a condition regarded as a disease. However, I contend that since all traits have genetic determinants and the "molecular dissection" of complex traits is increasingly possible, no characteristic that varies among humans is immune to genetic characterization as normal or abnormal. Although "drapetomania" appears to us to be a ludicrous disease ascription, the willingness of some slaves among the group to take the chance of trying to escape is not dissimilar to what contemporary molecular geneticists describe as "risk-taking" behaviour. Some research findings suggest a higher proportion of individuals who line up to try activities like bungee jumping have an allele "for" "risk-taking" behaviour than those who do not. Locating "mutations" in no way verifies or consolidates the disease status of an entity. Any values — biological, moral, aesthetic, social, or cultural — that attach at the level of the social and cultural being are imported to the level of the genome. Were a "gay gene" to be identified, it would represent an aberrant or disease allele to a society that regards homosexuality to be aberrant behaviour or disease; it would be just another

"normal" variant to a society where homosexuality falls within the limits of "normal" variation.

The polarized debate over whether disease is objective or culturally relative loses sight of what is most important to take from social constructivist accounts: that genotypes and traits alike have meaning for us only within *some* cultural context. This cultural contingency can be appreciated by cross-cultural comparisons. In western societies, treatment for infants born with CAH has typically been sex assignment, surgical intervention to correct ambiguous genitalia, and the institution of hormonal therapy and subsequent psychological counselling. With DNA testing now available by means of chorionic villus sampling (CVS) at eight to ten weeks of gestation, there are additional options. The hormone dexamethasone — a steroid which crosses the placental barrier to suppress the fetal adrenal gland — can be administered to the pregnant woman to coincide with the period of sex differentiation. Alternatively, a first trimester abortion can be carried out. Another pseudo-hermaphroditic condition is 5α-reductase deficiency. This condition is associated with an abnormally low conversion rate of testosterone to dihydrotestosterone (DHT), the hormone responsible for masculinizing the external genitalia of the male fetus. The consequence is ambiguous genitalia at birth. However, since pubertal sexual development is mediated by testosterone rather than DHT, virilization occurs. Testes enlarge and descend, there is enlargement of the "phallic clitoris," and erections occur with ejaculation through ducts in the lateral walls of the vagina or the urogenital sinus. Anne Fausto-Sterling (1992) draws attention to a small village in the Dominican Republic in which an American research team found that 5α-reductase deficiency occurs at a quite astounding rate — approximately one percent of all males. Affected individuals, recognized at birth due to ambiguous genitalia, are raised as girls until puberty whence the majority change their gender identity to male. Whereas in the Dominican community 5α-reductase deficiency resists the pathological label and is culturally accommodated, in western societies, any infant born with ambiguous genitalia becomes a patient, a candidate for immediate medical intervention, in fact, a "medical emergency."

To admit that clinical judgements of health and disease are informed by moral, aesthetic, social, and cultural, as well as biological, norms ought not lead us to throw up

our hands in despair at the impossibility ever of making reasoned decisions about appropriate practical interventions. Despite the inadequacies of functionalist accounts of health and disease and the intractability of debates between objectivist and social constructivist philosophers of medicine, it is important that discussions about health and disease continue. There needs to be a shift, however, from seeking universal and "objective" definitions for the concepts of health and disease (or for replacement notions such as "quality of life") to asking why *specific* conditions in a *specific* cultural context are considered diseases, dysfunctions, or disabilities. In each case, dominant cultural values that lie in the background must be brought to the foreground — what is implicit must be made explicit. In recognizing that judgements of health and disease are irreducibly normative, and always cultural as well as biological, we are forced to take responsibility for interventions that reinforce and perpetuate any moral, aesthetic, social, or cultural norms that are incorporated in such judgements. Biology alone cannot be the arbiter for difficult questions like whether we ought to "normalize" hermaphroditic bodies or change society to accommodate diversely sexed bodies.

4.4 Summary

In this chapter, I have argued that functionalist accounts of health and disease, and their extensions to "genetic medicine," are mistaken in two ways. First, following Canguilhem, in human molecular genetics as well as in physiology, the pathological is epistemically prior to the normal and not vice versa and clinical judgements of health and disease constitute distinctions between normal and abnormal biological functions and normal and mutant genes and not vice versa. Second, departing from Canguilhem, distinctions between health and disease and normal and abnormal biological functions are not "scientifically objective" or "weakly" normative, as I have defined these terms, but incorporate moral, aesthetic, social, and cultural norms. To the extent that our judgements of health and disease are value-laden, so too are those of genetic normality and genetic mutation. Genetic variation represents deviance, as well as deviation. But to recognize the cultural contingency of any standard of genetic normality ought not lead us to

foresake the possibilities of making reasoned and informed decisions about how to implement the new genetic technologies. The meanings of the concepts of health and disease and genetic normality and genetic mutation are negotiated and negotiable within culture. To justify clinical interventions by appealing to the presumed universality of an objective, theoretical, dispassionately rational standard is dangerous in several ways. For one, it operates to exclude some very central voices from the discussion — those who are immediately affected by disability — on the basis that their knowledge is "subjective." Another is that it leads us to expect that science and medicine will provide answers to problems that properly belong to everyone. And, finally, the values that tend to escape scrutiny in accounts that aim for universality are those most widely shared and dominant in a culture. As Evelyn Fox Keller (1992) writes: "All we have to fear today is our own complacency that there are some 'right hands' in which to invest this responsibility — above all, the responsibility for arbitrating normality" (p. 299).

Chapter Five

What's in a Cause?: The Pragmatic Dimensions of Genetic Explanations[1]

Hardly a week goes by in which we do not hear about a newly discovered gene for some condition or another. 'Geneticization' is a term used to describe this phenomenon marked by an increasing tendency to reduce human differences to genetic ones (Lipmann 1991). Traits which follow Mendelian patterns of inheritance have long been labelled 'genetic.' Mendelian or "single gene" diseases represent the success thus far of the "new genetics" — many such genes have been mapped, their mutations identified, and screening programs instituted. However, a rising number of more complex traits, ones for which environmental contributions are known to be significant, are also becoming viewed as 'genetic.' Heart disease, cancer, schizophrenia, crime, intelligence, and alcoholism are a few examples (Edlin 1987; Duster 1990). If we are to make sense of geneticization, it is necessary to understand the bases upon which traits are labelled 'genetic.' The ramifications of such designations are not inconsequential. One of the justifications for spending several billion dollars on human genome research is the belief that genes are key determinants of not only "single gene" but also complex traits. And insofar as theory directs action, 'genetic' problems call out for genetic (technological) solutions such that attention is turned toward individual bodies and away from our shared environments.

The preceding chapters have focused on distinctions between normal and abnormal genetic variation and normal and mutant genes. In this chapter, I am concerned with what it is to say that genes — whether normal or abnormal — cause traits. If there is one incontrovertible fact about genetic causation, it is that there are, strictly speaking, no "single gene" effects. *All* traits, no matter how simple, result from the interaction of many genes and the environment. It may seem trivial to assert that a trait can be deemed 'genetic' only relative to a necessary background of genetic and nongenetic factors. After all, not only is this accepted by biologists, but conditions we single out as causes, whether

[1] An article based on this chapter is forthcoming in *Biology and Philosophy*.

in science or in everyday life, are seldom sufficient for their effects. Causal claims foreground some factors and relegate others to the status of background conditions. However, at least since John Stuart Mill, much philosophical ink has been spilt over whether such distinctions are arbitrary, pragmatic, or objective. The question is not solely an abstract one concerning our ability to represent the causal structure of the world. Present-day efforts to explicate the notion of genetic causation and to define terms such as 'genetic trait,' 'genetic disease,' 'genetic basis,' 'genetic predisposition,' 'genetic susceptibility,' and 'genes for,' seek objective grounds for privileging genes as causes motivated by a belief philosophers and scientists share: that theoretical understanding furnishes the basis for rational action. The presumption is not only that one would intervene differently to prevent or to treat a disease depending on whether it is identified as 'genetic' or 'environmental' but that such practical concerns do not at all impinge on the scientific sphere.

In this chapter, I argue otherwise. Genes are singled out as causes within a practical, not theoretical, context. No trait can be said to be 'genetic' in an objective sense, if 'objective' is taken to exclude any pragmatic dimension. This is in no way to claim that genes are not "real" or that they lack causal efficacy. It is quite the opposite. Genes are singled out as causes not only because they are amenable to technological control but because they are perceived to be more tractable than their nongenetic counterparts and therefore the best means to various ends. By appreciating the pragmatic dimensions of genetic explanations, we come to understand the phenomenon of geneticization to be the consequence of an increased capacity to manipulate DNA in the laboratory and in the clinic, and not as an advancement in our theoretical understanding of "the way things really are."

In the first section of the chapter, I look at deterministic and probabilistic accounts of genetic causation, both those in which genes are purported to explain the presence of traits themselves and those in which genes (or, more properly, genetic differences) are understood to explain differences in traits. I emphasize three senses in which genetic explanations are context-dependent: first, genetic causes are singled out relative to a background of necessary genetic and nongenetic conditions; second, genetic explanations are population-specific; and third, genetic explanations are a function of the present state

of knowledge. In the second section of the chapter, I defend a pragmatic account of genetic explanation that draws on Bas van Fraassen (1980) and R. G. Collingwood (1938). I argue that how the cause-condition distinction is drawn, what population is selected, and which paths of research are followed, are practical choices that are influenced not only by theoretical considerations but also by the aims, interests, and orientations of those who make them. In the third section, I counter possible empirical and theoretical justifications for labelling traits and differences in traits 'genetic.' These include: causal priority, nonstandardness, and causal efficacy. In the fourth and final section of the chapter, I leave the question "why single out genes as causes?" to return to the problem of geneticization and the question "why this increased singling out of causes?" I argue that geneticization reflects, not an increased theoretical knowledge fostered by the development of DNA technologies, but the practical perception that such technologies have rendered genes easier to manipulate, and thus more convenient "handles," than environmental factors.

5.1 Deterministic and Probabilistic Looks at Causation

5.1.1 Explaining Traits

Twentieth-century philosophers have most often analyzed causation deterministically in terms of the logical interrelatedness of cause and effect — that is, whether a condition is necessary and/or sufficient for a given event to occur. Henrik R. Wulff (1984) and Philip Kitcher (1996) advance definitions of 'genetic disease' along these lines. Wulff emphasizes the necessity of the genetic condition. He distinguishes two senses of 'genetic disease': a strong sense applies where the genetic abnormality is both necessary and sufficient for the presence of the disease "regardless of environment"; a weak sense holds where the genetic abnormality is necessary but not sufficient, the disease occurring in most but not all environments. Kitcher, on the other hand, emphasizes the sufficiency of the genetic condition. He recognizes a strong sense of 'genetic disease' in which the presence of a mutation is associated with disease in "all known environments." In some of these cases, the disease is genetically determined, that is, the genetic condition is both

necessary and sufficient for its effect; in others, the genetic condition is merely sufficient since the disease may arise due to other (environmental) factors. Kitcher refers also to a weak, and overly liberal, sense of 'genetic disease' in which there is a "genetic basis" for the disease, meaning that it arises in some but not all environments given a particular genotype.

Since no gene or combination of genes is sufficient for a trait to develop, and genes are necessary for the development of all traits, the Wulff and Kitcher definitions of 'genetic disease' are forced to include *ceteris paribus* clauses that refer to an assumed background of necessary genetic and environmental factors. In this respect, the development of organismal traits, whether molecular, biochemical, or physiological, is no different from most other events, whether in science or in everyday life. Where there is no single condition that is both necessary and sufficient for an event to occur, we are nevertheless predisposed, it seems, to isolate one condition as "the" cause while ignoring the contributions of others. When we say that striking a match causes it to ignite, we tacitly assume the presence of oxygen and a dry match. Striking the match is not sufficient for it to ignite and neither is it necessary — holding an already lit match close by works even better. J. L. Mackie's (1965) solution is to introduce the notion of an INUS condition as the minimum requirement for designating causes. An INUS condition is "an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result" (p. 245) relative to a causal field of background conditions. Hence, necessary *ceteris paribus* content is made explicit: to single out an INUS condition as cause assumes that there is some basis for distinguishing it not only from the other necessary conditions that belong to its jointly sufficient set of conditions, but from the elements that belong to other sufficient sets of conditions as well as to the causal field.

To illustrate: take the example of phenylketonuria (PKU). PKU symptoms develop given the presence of phenylalanine in the diet and a mutation in the gene that codes for the enzyme that metabolizes the amino acid. PKU conforms to Wulff's weaker sense of a 'genetic disease': the mutation is necessary, and in most but not all environments sufficient, for the trait. For Kitcher, PKU is a genetic disease in that it has

a "genetic basis": given a particular genotype, it will arise in some, but not all, known environments. Mackie would say that the genetic mutation as well as a diet with phenylalanine are potential causes of PKU since both are INUS conditions belonging to the same sufficient set of conditions. For Wulff, Kitcher, and Mackie alike, if the gene mutation is to be singled out as the cause of PKU, this will be relative to a causal background that includes not only a diet comprising phenylalanine, but all of the genetic and nongenetic factors necessary for the development of an otherwise "properly functioning" individual. However, most diseases, even many Mendelian ones, are unlike PKU in that their presence in individuals cannot be predicted with near-certainty. Even though predictions based on DNA sequence are overwhelmingly statistical, very few philosophers have advanced accounts of genes as probabilistic causes. An exception is Kitcher's (1996) definition of the term 'genetic susceptibility' as an increased probability of disease in all known (strong sense) or some (weak sense) environments. An individual who possesses a particular mutation is, *ceteris paribus*, more likely to suffer the disease with which it is associated than someone who lacks the mutation. For example, a woman who inherits a copy of the BRCA1 gene has an increased chance of suffering breast cancer compared to one who has not.

5.1.2 Explaining Differences in Traits

Insofar as genetic and environmental factors alike play causal roles in the development of traits, the attachment of either a 'genetic' or 'environmental' label seems capricious. An alternate strategy is to emphasize that, while genes may not explain the development of traits in individual organisms other than by an arbitrary or pragmatic relegation of nongenetic factors to the causal background, genetic differences can be held responsible for phenotypic differences among members of a given population. Indeed, Richard Dawkins (1982) argues that differences in traits that develop via quite complex pathways (his example is reading ability) can be explained by single gene differences: "However complex a given state of the world may be, the *difference* between that state of the world and some alternative state of the world may be caused by something extremely simple" (p. 23). This is consistent with the approaches of evolutionary and classical geneticists who "blackbox" development in their respective studies of adaptation and heredity. It

seems a tall order to demand from such scientists, as "proper" explanations, ontogenetic ones. As a leading behavioural genetics textbook emphasizes, "the task of behavioral genetics is to determine the extent to which genetic variability accounts for behavioral differences among individuals" (Plomin, DeFries, and McLearn 1990, p. 104), and not to explain how behavioural traits develop in individuals.

Germund Hesslow (1983, 1984) adopts a populational approach to defining 'genetic trait' that is also deterministic. Hesslow argues that genetic explanations are always contextual. One never explains a property of an object *tout court* but only in relation to a reference class of an object or objects that lack the property. In effect, Hesslow's account incorporates all *ceteris paribus* assumptions into the explanandum, for these are the conditions shared by the object of investigation and the reference class object(s). 'Explanatory' causes are defined as the subset of 'determining' causes which the object of investigation does not share with the reference class object(s) and that were necessary for the actual effect and would have been sufficient for the effect in any member of the reference class.[2] According to Hesslow, which condition is singled out as cause depends on the composition of the reference class. He would say that in referring to PKU as 'genetic,' one explains why individuals with a specific genetic abnormality present with symptoms, while their reference class counterparts, who also ingest phenylalanine in their diet but lack the mutation, do not. However, should one wish to explain why symptoms are present in only one of two individuals, both of whom have the mutation, the "explanatory cause" will be environmental, that is, one's failure to follow the prescribed diet. Identically caused traits may be 'genetic' or 'environmental' depending on who is compared to whom. No trait is 'genetic' in any absolute sense, but only relative to a population.

Biologists similarly emphasize that to claim that a trait is 'genetic' is to do so relative to a specific population, and to the genetic and environmental factors shared by

---

[2] Gifford (1990) takes an approach similar to Hesslow's in seeking to describe the usage of the term 'genetic trait' in biology. He proposes two criteria, the major one of which, the "differentiating factor" (DF), is relative always to a particular population: "A trait is genetic (with respect to population P) if it is genetic factors which 'make the difference' between those individuals with the trait and the rest of population P" (p. 333).

members of that population. Their approaches, however, tend to be probabilistic, in cases both of qualitative and quantitative traits. For "qualitative" traits, Dawkins' (1982) account is much like Hesslow's, except that it is probabilistic rather than deterministic. For Hesslow, the presence or absence of a particular phenotype *corresponds* with the presence or absence of an identifiable genotype, relative to a reference class of individuals. For Dawkins, the presence or absence of a particular phenotype *correlates positively* with the presence or absence of an identifiable genotype, relative to a specified population and environment. Dawkins argues that terms like 'genetic trait,' 'genetic cause,' and 'a gene for' commit us only to the claim that, *ceteris paribus*, in some specified population and in some specified environment, a statistically more accurate prediction of phenotype is possible given knowledge of genotype. To say that there is 'a gene for' a trait, whether one as simple as eye colour in *Drosophila* or as complex as reading ability in humans, just means that, *ceteris paribus*, an individual with the gene is more likely to possess the trait than one without the gene.

Qualitative variation in a trait in a population may be associated with variation in more than one causal factor. For example, epidemiological studies report positive correlations between cardiovascular disease and factors such as smoking, heredity, stress, obesity, and high cholesterol levels. While recognizing that both genetic and environmental factors contribute to complex traits like cardiovascular disease, molecular geneticists seek to identify single gene loci of causal significance through segregation and linkage analysis. The identification of "positive (or negative) causal factors" at the level of populations is consistent with the probabilistic accounts of causation proposed by many philosophers of science which serve to update Carl Hempel's model of inductive-statistical (I-S) explanation. The basic idea is that statistical relevance — a difference between conditional and unconditional or posterior and prior probabilities — is a better indication of a causal relation than are high probabilities (Suppes 1970, Salmon 1984). Wesley C. Salmon's (1984) approach is to identify those statistically relevant factors which partition a reference class of individuals into a number of homogeneous cells each with a different likelihood of sustaining a given effect. For Elliott Sober (1984a), a "causal factor must

raise the probability of the effect in at least one background context and must not lower it in any" (p. 294).

Quantitative geneticists are concerned with traits such as height, weight, intelligence, or fecundity, that vary continuously between individuals and are therefore likely to be modulated by multiple causal factors, both genetic and environmental. Recently, molecular techniques have been developed to identify quantitative trait loci, known as QTLs (Plomin, DeFries, and McCleam 1990; Kearsey and Pooni 1996). This is done by screening polymorphic markers, singly and in combination, for their contributions to variance in quantitative traits. The traditional quantitative genetics approach uses statistical methods such as analysis of variance (ANOVA), analysis of covariance (ANCOVA), and path analysis to estimate the relative contributions of genes and environment to continuously varying traits. ANOVA of suitable groups can partition phenotypic variance into additive and nonadditive variance. Additive components include genetic and environmental variance; nonadditive components include dominance, epistasis (gene-gene interaction), gene-environment covariance, and gene-environment interaction. Heritability expresses the percentage of total phenotypic variance that is due to genetic variance: "broad-sense heritability" comprises both additive and nonadditive genetic variance; "narrow-sense heritability" includes additive genetic variance only. Any heritability coefficient is relative to a specific population, its value dependent on the particular distribution of genotypes and environments in this population. A trait's heritability will be zero, not only where genetic variation does not contribute to phenotypic variation in the trait, but where the trait itself or the genes that contribute to the trait do not vary in the population.

On population-based accounts of genetic causation, where genetic differences are understood to explain differences in traits and not traits themselves, it seems possible to escape the problem of arbitrarily or pragmatically singling out some necessary conditions over others as causes. Causes are differences among members of the population that are associated with variation of the trait in question. Background conditions are either shared by members of the population or, where they differ, do not contribute to variation of the trait. The result, of course, is that no trait is 'genetic' except relative to a particular population. The same trait may be considered to be 'genetic' in one population and

'environmental' in another. For example, a disease such as lactose intolerance is considered to be a genetic disease in Northern European populations where ingestion of milk products is common and lactase deficiency rare. In African populations, where ingestion of milk products is rare and lactase deficiency common, it is considered an environmental disease (Hesslow 1984, p. 189). It is the composition of the reference class or population that determines which factors are singled out as causes. Therefore, the choice of reference class or population requires an objective, nonarbitrary and nonpragmatic, basis.

5.1.3 Genetic Causation

There are several reasons to prefer a probabilistic account of genetic causation to a deterministic one. Even relative to a background of those genetic and nongenetic conditions consistent with the organism's "normal" development and function, there is seldom one-one correspondence between genes and traits. Epistasis (the interaction of many genes to produce a single trait) and pleiotropy (the effects on many traits by a single gene) are widespread. The probabilistic approach also more effectively manages the genetic, causal, and phenotypic heterogeneity that presents problems for the one-one mapping of causes and effects, or genes and traits, deterministic accounts expect and require.[3] Adopting the probabilistic approach not only accommodates these facts but it prevents the reification of the distinction between simple Mendelian and complex non-Mendelian, and qualitative and quantitative, traits that occurs when the first of each pair is treated deterministically and the second probabilistically. Those traits considered to be "genetically determined," since they arise in all environments given the presence of a particular genotype and the necessary genetic and environmental causal background, serve as limiting cases where posterior probabilities of the trait given the gene are close to 1.

---

[3] Genetic heterogeneity is of two sorts: a particular set of symptoms classified as a single disease may be associated with mutations of different alleles or with different mutations of the same allele. Causal heterogeneity refers to the fact that the same disease may arise for either predominantly genetic or predominantly environmental reasons. Phenotypic heterogeneity refers to the wide range of symptoms that may be associated with identical mutations, even in "single gene" diseases; geneticists also use the terms 'penetrance' and 'expressivity' to refer to such phenotypic variability.

Also advantageous is the probabilistic approach's explicit recognition that reference class partitions depend both on the causal structure of the world and our knowledge of it; the homogeneity of cells may be either objective or epistemic. This better accommodates the fact that traits are "genetically determined" relative to *known* environments.

However, my main concern in outlining both deterministic and probabilistic approaches to genetic causation is to point not to their differences but to an important similarity they share. Genetic explanations, whether understood deterministically or probabilistically, are context-dependent. To label a trait 'genetic' is to make a claim that is relative in three ways. First, genes can be considered causally efficacious only relative to a causal background of also necessary genetic and nongenetic conditions. Second, isolating genes as causes is a function of the composition of a particular reference class or population. Third, singling out genes as causes is contingently dependent on the current state of knowledge and does not necessarily reflect the causal structure of the world. This is true whether a reference class is effectively partitioned by a single genetic factor (as for so-called "genetically determined" traits where genes are both necessary and sufficient conditions relative to a causal background) or by multiple genetic and nongenetic factors (as for so-called complex or "multifactorial" traits where genes are neither necessary nor sufficient conditions relative to a causal background) in cases of qualitative traits. It is true, also, for estimates of heritability in cases of quantitative traits. In the following section, I argue that pragmatic factors are influential in determining such choices: the distinction between causes and conditions; the target population or reference class; and research directions, hence, the state of knowledge concerning various traits.

It might be argued that, by treating genes as explanatory of differences in traits and not of traits themselves, we can forego worrying, at least directly, about the possibility that one cannot objectively distinguish between causes and conditions. The distinction would merely be the indirect consequence of the properties of the population one chooses. This would leave us only to worry about genetic explanations being population-specific and epistemically relative. However, at least two factors mitigate against this. First, it is believed that genes do "directly cause" at least some traits. Dawkins (1982), a prominent biologist and advocate of the populational approach, explicitly singles out protein *production* as an example of genes determining traits

themselves, and not just differences in traits: "Other than at the molecular level, where one gene is seen directly to produce one protein chain, geneticists never deal with units of phenotype as such. Rather, they always deal with *differences*" (p. 21). But, even in protein synthesis, genes are *not* sufficient for their effects. Genes alone do not make peptides, proteins, gross organismal traits, or organisms themselves. The environment, considered in its widest sense to include all epigenetic factors, internal as well as external to the organism, plays a causal role in protein synthesis — through the availability of the necessary amino acids, enzymes, temperature, etc. In the words of Kelly C. Smith (1992): "Only a complex *system* containing genes as one of its (necessary but insufficient) components is capable of protein production" (p. 338). This is one reason to examine possible objective bases for privileging genes over nongenetic conditions in the determination of traits themselves, and not just differences in traits.

The other reason is that it is not at all clear how well identifying causes of phenotypic differences among members of a population serves to provide theoretical justification for how best to intervene in individuals. And, after all, this is why philosophers are preoccupied with the identification of objective (nonarbitrary and nonpragmatic) bases for calling certain traits 'genetic.' Explaining trait differences in terms of genetic differences may suffice for predictive or diagnostic purposes where intervention takes the form of abortion or embryo selection, or mate selection. However, should the desire be to intervene in an existing individual, the 'genetic' label in no way legitimizes that this take place at the level of DNA — other forms of intervention may be equally or even more effective. Take, for example, an individual with PKU or lactose intolerance. For the purposes of treatment, there is no reason to discriminate between someone for whom the disease is 'genetic' relative to a reference class of individual(s) who lack the relevant mutation and ingest phenylalanine or lactose and someone for whom the disease is 'environmental' relative to a reference class of individual(s) who share the relevant mutation but do not ingest phenylalanine or lactose. In both cases, manipulation of the genome or the diet may prove effective. What matters is the interaction of causal factors in the development of disease symptoms in individuals, and not the source of differences between individuals. This means that, if the goal in labelling traits 'genetic' is to provide objective grounds to intervene, we cannot escape the

problems posed in the privileging of genes as causes in the genesis of traits by stipulating that genes explain differences in traits and not traits themselves.

5.2 The Pragmatic Dimensions of Genetic Explanations

Philosophers who seek objective (nonarbitrary, nonpragmatic) bases for labelling traits 'genetic' are often interested in providing theoretical justifications for practical interventions. In this section, I argue that genetic explanations are pragmatic, or, in other words, that practical, not theoretical, considerations direct the singling out of genes as causes. This means that explanatory context matters. In the case of genetic explanations, there is a plurality of relevant explanatory contexts — scientific, experimental, clinical, social, and economic — and each is shaped by various interests.

R. G. Collingwood (1938) and Bas van Fraassen (1980) argue that causal explanations, unlike theoretical structures, are context-dependent. Collingwood relies on a distinction between the theoretical and "practical natural sciences," contrasting the notion of 'cause' as it appears in each. Theoretical causes are necessary; they are replaceable by reference to laws and their instances. Where control over nature is sought, on the other hand, causal language expresses the finding of "certain means useful to certain ends": "In this [practical] sense, the 'cause' of an event in nature is the handle, so to speak, by which we can manipulate it" (p. 89). Hence, for Collingwood, practical causes are contingent in two ways: one, they are relative to specific human purposes; and two, their effects are fulfilled only in combination with other necessary conditions. Van Fraassen's contention in *The Scientific Image* that explanations conform directly to the aims, interests, and orientations of different scientists or groups of scientists and only indirectly to the nature of the phenomena rests on his distinction between theory and explanation. According to van Fraassen, theories are context-independent; they describe or "save" the phenomena and permit accurate predictions to be made. Explanations, on the other hand, are context-dependent; they are responses to "why-questions" explanation-seekers find meaningful. Although there is a "causal net," defined by van Fraassen as "whatever structure of relations science describes," "which could *in principle* be described

in detail," causal explanations are partial insofar as they draw attention only to "certain features of the causal net" (pp. 124-125).

Insofar as explanations are answers to "why-questions," van Fraassen argues that pragmatic considerations structure the questions that are asked, and, in so doing, determine the range of possible responses. Questions assume a relevance relation: our aims, interests, and orientations dictate whether we seek information about causal mechanisms, goals, desires, etc.[4] As experimentalists, molecular biologists seek control over nature and the phenomena they generate in their laboratories. The aims of technological control associated with laboratory research lead to questions that assume a causal relevance relation which is linear and unidirectional, since experimentation proceeds by measuring perturbations in a system subsequent to the isolation and manipulation of its parts. This mechanistic conception of the experimental system is consistent with, at least methodological, commitments to determinism and reductionism. Abraham Kaplan (1965) goes so far as to suggest that causal concepts arise in science out of the desire to exert technological control over nature and that the cause-condition distinction makes sense only "by reference to the possibility, or desirability, of intervention by the experimenter" (p. 147). In other words, identifying some necessary conditions — say genes — as causes, and relegating others — say environmental factors — to the causal background, is a product of the laboratory and a function of the technological power to intervene. What we isolate as "the cause" is, in Collingwood's words, the "means" most useful to "certain ends" or the "handle" by which it is possible to "manipulate" an outcome.

The discovery of recombinant DNA technologies in the early 1970s permitted a precision of experimental control over the hereditary material that was not possible with radiation-induced mutation. Genes became manipulable and, as Kaplan would have it, bonafide causes. Certainly, given the interventionist aims of a laboratory science, DNA is less unwieldy than large-scale entities of causal significance like electromagnetic

---

[4] Aristotle's typology of four causes is a good example of different relevance relations. Also familiar to biologists and philosophers of biology is Mayr's (1961) distinction between the proximate causation sought by functional biologists such as physiologists, functional anatomists, and molecular biologists, and the ultimate causation sought by evolutionary biologists.

radiation, air pollution, or poverty. However, many other nongenetic factors both internal and external to the organism are amenable to experimental manipulation. Why focus on genes? Van Fraassen argues that pragmatic aims determine not only the relevance relation but, both directly and indirectly, the specific way in which the cause-condition distinction is drawn. It is impossible to describe the entire causal net that surrounds an event. That we find some factors rather than others salient, and identify these as causes, is a direct reflection of our particular set of interests. Who asks questions matters. Only certain questions are likely to arise within a group that shares a particular set of theoretical and factual commitments. Just as the lawyer, engineer, and mechanic are likely to identify different causes of a motor vehicle accident, so might the surgeon, epidemiologist, and molecular geneticist focus on different factors associated with disease. Questions about genes as causes tend to arise in molecular biology because of its commitments to Weismannism and the Central Dogma and its acceptance of metaphors that characterize DNA as "master molecule" or the "program that computes the organism." Focusing on genes makes perfect sense if assumptions that only germ cell nuclei are inherited from one generation to the next and that information flows unidirectionally from nucleic acid to protein are unquestioned. Along these lines, Evelyn Fox Keller (1992) argues that the concept of 'genetic disease' has been extended beyond what can be justified empirically due to biologists' beliefs in genetic determinism. Similarly, Ruth Hubbard (1990, with Elijah Wald 1993) contends that scientists focus on genetic causes while ignoring environmental ones because of their commitment to a reductionist approach that attempts to explain the functioning of a complex system *only* in terms of its smallest constituent parts.

According to van Fraassen, pragmatic goals also indirectly influence how the cause-condition distinction is drawn because they determine the contrast-class of propositions that delimits the topic of the question and therefore the appropriate range of responses. Recall the several different meanings van Fraassen's well-known example "Why did Adam eat the apple?" can have depending on what we understand the contrast-

class to include.[5] Since genes exert their effects only in the presence of other genes and environmental factors, singling them out as causes requires the enclosure of this necessary backdrop within an assumed *ceteris paribus* clause, the content of which is fixed by that of the contrast-class. However, in no way do contrast-classes unwittingly impose *ceteris paribus* content upon us; they are chosen pragmatically, in conformity with *ceteris paribus* assumptions that fulfill specific theoretical, methodological, and experimental aims. Choices are made about what to vary and what to keep constant, what to foreground and what to relegate to the background. By electing to control for environmental factors in the laboratory, genes are rendered the target of causal investigation.

Given that explanations are contextually determined by the aims, interests, and orientations of those who seek them, it is hardly surprising that any number of conditions might be selected as "the cause" of a given event. Nor is it surprising that where aims, interests, and orientations intersect, similar explanations will be found. Following on Kaplan's words, the focus on genetic causes reflects not only the "possibility" born of technological innovation, but the "desirability" born of professional interest. Molecular geneticists obviously have a professional stake in maintaining a focus on the causal efficacy of genes. The rise of molecular genetics as a subdiscipline of molecular biology was furthered by the development of recombinant genetic technologies and the privileging of DNA by the theories and metaphors of molecular biology. However, it is interesting to note that in areas of molecular biology other than molecular genetics there is movement away from the Central Dogma's simple linearly causal model and toward an appreciation of the organism as a complex system of which genes are necessary, but not the sole, components (Keller 1994, 1995; Strohman 1993). The professional self-interest of molecular geneticists is only part of the reason for this divergence. The other is the clinical context in which, particularly human, molecular genetics research is carried out.

---

[5] We understand the question to mean 'Why was it *Adam* who ate the apple?,' when we take the contrast-class to include others who were present such as Eve or the serpent; to mean 'Why was it the *apple* Adam ate?,' as opposed to a pear or some other fruit; to mean 'Why did Adam *eat* the apple?,' when the contrast-class includes other possible options such as throwing or stepping on the apple.

Here, the goal of intervention is not just experimentation, but the prevention, treatment, and eradication of disease. As Collingwood notes, causes are singled out from among other necessary conditions according to a practical criterion of "what can be put right." Richard C. Strohman (1993) observes that researchers in biomedicine and behaviour, the "applied medical sciences," maintain fast hold on the "genetic paradigm" that "basic research biology" is abandoning (p. 117). If "producing or preventing" one factor "puts things right," why worry about the others? That it is genes that are singled out as the most tractable factors in clinical research fits well with the traditional North American approach to medicine which, in its assumption of a biological and reductionist model of disease, focuses on internal, rather than external, factors in pathogenesis.

However, the practical context for the "molecularization" and geneticization of disease extends far beyond making individual patients better and furthering disciplinarian professional interests. It incorporates the wider social and economic interests of molecular geneticists, other investors in the biotechnology industry, university and private patent-holders, and government. As legal scholar Philippe Ducor writes: "The advent of biotechnology has ... virtually eliminat[ed] the traditional distinction between 'basic' and 'applied' research" (p. 13). Patents can be held on stretches of DNA that prove to have applications; hence, financial returns will depend ultimately on successful marketing of applications such as screening tests, genetically engineered pharmaceuticals, gene "therapies," etc. Many molecular biologists have economic interests in biotechnology companies — as owners, directors, or shareholders. This intersection of interests began with the invention of recombinant DNA technologies and the ability to generate huge quantities of human proteins using bacterial cloning for commercial retail. Only in the past five years has genomics — the construction of genetic and physical maps, gene mapping, and genome sequencing — seemed a viable investment for venture capitalists and the large pharmaceutical companies. "[G]enome research has become a veritable hotbed of capitalism" (Anderson 1993, p. 300), wrote one science reporter in 1993. He noted that thirty leading genome scientists had made commercial deals with venture capitalists. These scientists included Leroy Hood and James D. Watson in the area of high-speed sequencing technology, Craig Venter in the area of cDNA sequencing, and Daniel Cohen, Walter Gilbert, and Eric Lander in the areas of disease gene identification

and the development of gene-based therapeutics. Shortly after this, genomics companies began to attract financial backing from pharmaceutical giants looking for a new investment course after the industry's downturn in the early 1990s due to the failure of conventional in-house research programs to produce enough new medicines (Abelson 1996). In 1995, pharmaceutical companies spent $3.5 billion to acquire biotechnology companies, $1.6 billion on research and development licensing agreements, and $700 million to obtain access to genome databases maintained by biotechnology companies (ibid.). For example: pharmaceutical giant Eli Lilly is backing a commercial venture by Gilbert and Mark Skolnick to develop gene-based cancer therapies (Anderson 1993); Novartis, the second largest pharmaceutical company in 1996 following a merger between pharmaceutical giants Ciba-Geigy and Sandoz, contributed one million dollars to Lander's gene-mapping efforts at Whitehead Institute-MIT Center for Genome Research (Koenig 1996); also at Whitehead-MIT, a consortium of companies led by pharmaceutical giant Bristol-Myers Squibb has invested $40 million in a 5-year initiative to find more efficient ways to gather and to compare genetic data (Roush 1997). Although this intersection of knowledge and profit may suggest troublesome conflicts of interest, it should be noted that the United States government has been motivated to fund the Human Genome Project for the sake of the health, not only of the American people, but of its developing biotechnology industry.

5.3 Possible Objective Criteria for Singling Out Genes as Causes

Although numerous empiricist philosophers have followed Mill in holding that the cause-condition distinction is invariably an arbitrary or pragmatic one, this is by no means a majority view. Many criteria have been advanced for objectively selecting causes from among possible contending conditions. While the diversity of these criteria (manipulability, frequency of occurrence, irreplaceability, nonstandardness, causal efficacy, blameworthiness, causal priority) is taken by van Fraassen to indicate the context-dependence of what we take to be explanatorily relevant and the wide variety of interests that motivate us to seek explanations, these claims nevertheless need to be

evaluated. Some of the criteria are relevant to genetic causation: they purport to offer nonarbitrary, nonpragmatic grounds for distinguishing between the genetic condition singled out as the cause and the genetic and nongenetic conditions relegated to the necessary causal background. I consider three such criteria in this section: causal priority, nonstandardness, and causal efficacy. Causal priority can be conceived solely in terms of the explanation of traits in individuals; nonstandardness applies both to populations differentiated by a single causal factor and to individuals who are members of such populations; and, causal efficacy is a consideration for the explanation of differences in traits in populations only.

5.3.1 Causal Priority

There are several ways in which causal priority may be established. First, the causally prior condition is sometimes identified as the one that initiates chains of events that occur *within* the body (Nordenfelt 1981). From the standpoint of theory, however, it is completely arbitrary to restrict causal chains leading up to some end event to those that lie inside the body. Take, for example, the "two-hit" hypothesis that dominates cancer research. It explicitly recognizes the roles of both heredity and the environment in predisposing individuals to cancer. Nevertheless, even though environmental carcinogens are recognized by researchers to be causally efficacious agents associated with mutations in somatic cell DNA, the boundary between body and external environment is used to focus attention on these DNA changes as the "foundation" and "starting point" for cancer: "If we had not been able to study cancer at the level of the change in DNA that starts it, the disease would still be a hopeless field.... Not until the genetic foundation for cancer was identified could you really begin to say what goes wrong to make this terrible human affliction" (Watson 1992, p. 166). Cancer is now considered by some prominent molecular biologists, whether associated with inherited mutation or not, to be a genetic disease — even when necessary, and prior, environmental causes are recognized: "Cancer, scientists have discovered, is a genetic condition in which cells spread uncontrollably, and cigarette smoke contains chemicals which stimulate those molecular changes" (Bodmer and McKie 1994, p. 89). Although there is no theoretical basis for restricting causal candidacy to conditions that lie within the body, this approach does

fulfill a couple of pragmatic aims. One is that *some* cause must be singled out, the potentially infinite regress of possible causes arrested, if there is to be a "handle" by means of which to intervene. Another is that treating the body as a closed system serves spatially to confine potential causes of disease so that they can be more readily localized. This not only ensures a "handle" which is convenient, but represents, from the outset, a practical commitment to disregard interventions that address the individual's physical and social environment.

From the restriction of the causally prior condition to that which lies inside the body, it need not follow that this condition be genetic. Genetic causes are privileged because molecular biology's central tenets accord DNA both temporal and ontological — and, hence, causal — priority. By temporal priority, I mean that DNA (as at least a near proxy for what we mean by 'gene') in some sense exists before other cellular components and the organism itself. Weismannism accords DNA temporal priority as a physical entity that is present, at least as part of what will become the embryo, before all the other physical elements also necessary for development to ensue. This is because the doctrine assumes that only germ cell nuclei are continuous from generation to generation, and that somatic cells and germ cell cytoplasm are discontinuous across the generations, arising anew in each. The Central Dogma of molecular biology represents a 1950s reformulation of Weismannism in terms of information theory. It asserts that information travels unidirectionally from nucleic acids to protein, and never vice versa. Here, DNA is no longer temporally prior in the physical sense; rather, it is the point of origin for the transfer of information. The chief difficulty with the Weismannism-Central Dogma claims of temporal priority, whether physical or informational, is of the chicken-and-egg variety. It is questionable whether DNA can be considered to be temporally prior to other molecules in the body's internal milieu. As Smith (1992) argues, since nucleic acids need proteins and other cellular components to make proteins, DNA cannot be accorded temporal priority whether we are attempting to explain the origin of life on earth or embryogenesis (the development of the individual). While theories about life's origins remain entirely conjectural, it is fully accepted that the fertilized ovum contains the cytoplasmic contribution of at least the maternal germ cell. Yet, there persists, in

developmental genetics, a tendency to focus on cytoplasmic (mitochondrial) DNA and to ignore the role of cytoplasmic proteins.

Finally, the Central Dogma, along with other informational metaphors so prevalent in molecular biology such as the genetic "code," and DNA as "master molecule" or the "program that computes the organism," attribute ontological priority to DNA. By ontological priority, I mean the privileging of DNA over other molecules based on its essential nature — as per James D. Watson's description of DNA as "the most golden of molecules" (in Bodmer and McKie, p. 10). There are good reasons to be sceptical of appeals to "information" in molecular biology that attribute ontological priority to DNA. There is, undeniably, a formal relationship between the sequence of amino acids in a polypeptide or protein molecule and the sequence of nucleotides in the segment of DNA that "codes" for it. We might look at this in two different ways. The first is to treat DNA sequence data as the axiomatic foundation of a deductive structure. This represents Walter Gilbert's (1992) dream to achieve one day a fully theoretical biology in which accurate predictions about the linear and three dimensional configuration of proteins, protein function, and the structure and function of the organism as a whole would flow from knowledge of DNA sequence alone. Of course, while it seems unlikely, should this happen, there would be no longer any need to talk about causation. One could just as readily predict DNA sequence from the level of protein or organism, as vice versa. However, even at the lowest levels of organization, Gilbert's dream faces formidable obstacles. Sahotra Sarkar (1996) points out that, notwithstanding the protein folding problem and the need to consider gene regulation in order to proceed beyond the level of protein structure, one faces significant difficulties in attempting even to predict the linear structure of proteins from sequence data alone: specifically, the ability to recognize transcription initiation sites and, in the presence of extensive RNA editing, the boundaries between introns and exons and coding and noncoding segments of DNA. Sarkar concludes: "the code ... is of little predictive value in novel contexts" (p. 201).

This relationship between DNA and protein sequences can also be regarded in a second way, in terms of Aristotle's notion of a formal cause. What Francis Crick meant by information was "the specification of the amino acid sequence of the protein" (in Sarkar, p. 196), and the comparison of the DNA "code" to the idea of the statue's form

that precedes and guides the artisan's sculpting of the statue seems a good one. As opposed to a fully deductive biology, here, the relationship between DNA and protein sequences is one of causal asymmetry. Provided that all of the cell components necessary for protein synthesis are present, modification of the DNA sequence may be followed by a predictable and specifiable change in protein sequence. The opposite will not occur. Fred Gifford (1990) conveys this idea in his "proper individuation (PI)" criterion for a definition of 'genetic trait': "For a trait to be genetic, the gene (or set of genes) must cause that trait *as described.* The trait must be individuated in such a way that it matches what some genetic factors cause specifically" (p. 343). Gifford uses protein structure as an example of a trait that "genetic factors cause specifically." Smith is correct in his response to Gifford that "[o]nly a complex *system* containing genes as one of its (necessary but insufficient) components is capable of protein production (and thus of structural and catalytic activity)" (p. 338). Nevertheless, this ignores the distinction between formal and efficient causation. It is legitimate to argue that, while there is no nonpragmatic or nonarbitrary way to single out genes as efficient causes of protein synthesis over other, also necessary, cellular components, gene sequences are uniquely formal causes.

Even if we accept the notion of formal causation (and only in the restricted sense in which protein sequence is determined by DNA sequence and not vice versa), this need not privilege the causal contributions of genes or DNA. Informational metaphors have accorded genes, unlike other cellular components, directive agency, as illustrated in this quote from Watson: "Ignoring genes is like trying to solve a murder without finding the murderer. All we have are victims" (1992, p. 167). Keller (1995) notes the "two-sided image of the gene, part physicist's atom and part Platonic soul" (p. xv) that has persisted since Schrödinger's "'law-code and executive power—...architect's plan and builder's craft—in one'" (p. xv), resulting in a "discourse of gene action" that attributes "agency, autonomy, and causal primacy to genes" (p. 8). However, contingent historical and social forces lie behind this "causal primacy of genes." First, and foremost, it is supported by a metaphysical preference for form over matter and mind over body that has more than 2000 years of history in western civilizations. R. C. Lewontin (1993) suggests that it is also a more contemporary manifestation of capitalist ideology — minds elevated to the

boardroom and bodies confined to the factory floor below. Keller (1995) cites several other social and political factors that contributed to the privileging of DNA over protein and nucleus over cytoplasm in the first half of this century: the gap between genetics and embryology, the identification of the nucleus with American interests and the cytoplasm with European, especially German, interests, and the treatment of the nucleus as male and the cytoplasm as female.[6]

5.3.2 Nonstandardness

For qualitative traits, nonstandardness is proposed as an objective criterion for designating a trait 'genetic' where genetic factors either provide necessary and/or sufficient conditions for the trait's development given the necessary genetic and nongenetic background conditions or increase the likelihood that the trait will arise given these background conditions. Causes and background conditions are distinguished on the basis of what is abnormal or normal, unusual or usual, or nonstanding or standing. Recall that Wulff and Kitcher take this approach in defining 'genetic disease': Wulff's definitions refer to "most or all environments"; Kitcher's robust definition refers to "all known environments." In other words, in a given individual, a disease is 'genetic' if genetic factors are necessary and/or sufficient for the disease given the presence of *normally* occurring genetic and nongenetic factors. Unlike the criterion of causal priority, the criterion of nonstandardness applies to the presence of traits in individuals only as members of a population or in comparison to a reference class. What is taken to be nonstandard or standard, and therefore foregrounded as cause or relegated to the causal

---

[6] I argue here that it is legitimate to understand the specification of protein sequence by DNA (exon) sequence in terms of formal causation but that from this need not follow the privileging of genes as causes. Some contend that we ought to dispense with informational metaphors and formal causes altogether. For instance, Sarkar (1996) suggests that the concept of biological specificity in Linus Pauling's sense in which the shape of a molecule determines its behaviour be resurrected and updated to obviate the need for informational metaphors that Schrödinger's focus on the arrangement of a molecule's units requires. In place of genetic reductionism, one could achieve a "thoroughly physicalist reductionist account of the interactions between DNA, RNA and protein" (p. 218) in which "[c]oding will be retained only as a short-hand description of the usual triplet specification of amino acid residues, but it will not be assumed to have any explanatory value" (p. 222). If Sarkar is right, then, even if legitimate, it might be unnecessary to treat DNA sequence as the formal cause of protein sequence.

background as the case may be, depends on the frequency of properties across a number of individuals. A trait is 'genetic' only relative to a specific population or reference class and the properties that belong to members of that population or reference class.

The choice of population or reference class therefore determines how the cause-condition distinction is drawn. Kim Sterelny and Philip Kitcher (1988) formalize Dawkins' probabilistic explication of the concept of 'a gene for' in terms of the "standard" environment of genetic and nongenetic factors shared by members of a population. "Nonstandard" genetic and nongenetic factors are ruled to be those factors which are unlikely or infrequent, or would, more precisely, interfere with development so as to preclude any expression of the trait the allele in question is said to cause. Individuals are excluded from membership in the population should any such "nonstandard" properties belong to them. Hesslow similarly emphasizes that the choice of reference class is neither arbitrary nor pragmatic, but objective. It is the fact that certain regularities exist — what is empirically normal or standard or what is theoretically or morally ideal — that permits us to secure the "true explanandum," in most cases these "regularly" occurring objects being the obvious candidates for reference class membership. The causal background of "normally" occurring genetic and nongenetic factors that is incorporated into the explanandum includes those factors the object of investigation shares with reference class objects. Hence, the cause-condition distinction is not at all arbitrary: "explanatory" causes explain the locally abnormal, the locally unusual, and the locally deviant.

Van Fraassen's contrast-class is structurally similar to Hesslow's reference class. Both van Fraassen and Hesslow contend that explanations are contextually dependent on the choice of contrast-class or reference class. Context is provided for van Fraassen by the propositional content of the contrast-class and for Hesslow by the properties of reference class objects. Counter to Hesslow, however, van Fraassen argues that the choice of contrast-class is pragmatic, not objective. If causes are singled out because they are "abnormal" or "unusual," this merely reflects a particular set of practical explanatory aims. In this way, van Fraassen's pragmatic account better accommodates the experimental context in which molecular genetics research is carried out. Hesslow's approach, as well as that of Sterelny and Kitcher, assumes that populations and reference classes thrust

themselves upon passive observers of nature who exercise no choice regarding what is to be explained relative to what. In experimental biology, practical decisions are involved in choosing the subject population of organisms (or cells, etc.), as well as the environmental conditions under which they are to be studied. Experimental geneticists choose what to vary and what to keep constant depending on their aims, interests, and orientations: whether to induce genetic variation through breeding, radiation-induced mutation, or gene insertions and "knock-outs,' or to manipulate nongenetic factors either internal or external to the organism. Quantitative geneticists make similar decisions in selective breeding and crossing experiments. Experimentalists are usually well aware of the context-dependence of their research and the need to question its applicability to the world outside of idealized laboratory conditions. However, since causal knowledge is often discovered through experiment by varying laboratory conditions beyond the limits "normally" found in the world, what is "normal" or "standard" cannot provide the objective basis for foregrounding some (genetic) conditions and relegating other (nongenetic) ones to the causal background.

It is easy to see how pragmatic aims guide experimental research; after all, in such settings, practical decisions about how to intervene are inevitable. Outside the laboratory, however, nonstandardness represents a serious challenge to a pragmatic account of genetic explanation. In cases like Huntington's disease or sickle cell hemoglobin, a single genetic mutation is necessary, and arguably sufficient given necessary (and standard) background conditions, for the trait's development. While counter-examples like the individual who dies in a motor vehicle accident prior to the onset of Huntington's, or the excessive temperatures that preclude the synthesis of hemoglobin polypeptides of any sort, are surely stretchings of a logical point, their value is to emphasize that a sequence of events, to which nongenetic factors are necessary contributers, precedes the appearance of any recognizable phenotype. To assert that a trait is "genetically determined" is to make an epistemically relative claim that, given the presence of the genetic abnormality, the trait appears in all *known* environments. In probabilistic terms, one says that the homogeneity of the cells created by the effective partitioning of any possible reference class by the genetic factor is epistemic, not objective. There may exist an alternate genetic and/or environmental background —

whether to be discovered or created — in which the trait will not appear. The removal of phenyalanine from the diet of those born with the mutant gene associated with PKU provides such an example.

However, attaining knowledge of the mechanisms by which genes and environment interact in the production of disease, so to make possible alternate means of intervention, is a function less of nature's intrinsic properties than of pragmatic choices about how research efforts ought to be expended. In human molecular genetics, such choices are influenced by clinical and social, as well as scientific, contexts. Research into individual diseases tends to fall by the wayside as soon as a socially accepted means for their control is attained. Consider Down's syndrome. Trisomy-21 is an infrequently occurring genetic condition that is both necessary and sufficient for Down's to occur, given a standard causal background of genetic and environmental conditions that excludes any conditions incompatible with otherwise "normal" development. It seems, given what appears to be an entirely objective basis for designating the trait 'genetic,' there is no room left for arbitrary or pragmatic choices. But this is not the case. Since accurate predictions of Down's can be made on the basis of prenatal tests, if abortion is an acceptable means of intervention and there is widespread agreement that individuals who are affected to any degree ought not knowingly to be brought into the world, there is no practical incentive to continue research into the condition. Identifying the chromosomal abnormality as "the" cause suffices. But if some prospective parents want to know what the nature and severity of symptoms are likely to be in order to decide whether to abort the fetus or bring it to term, the explanandum changes, from that of Down's syndrome *simpliciter* to the nature and severity of Down's-related symptoms. The presence of an extra chromosome-21 no longer serves as a suitable explanation. Additional genetic and nongenetic causal factors will need to be identified to make accurate predictions of this order.[7] Research direction may even be motivated by an individual scientist's moral

---

[7] Clinicians face difficulties, even in the most researched of "single gene" diseases, in predicting the form, severity, and age of onset of symptoms from knowledge of the genetic defect alone. Although research efforts are being directed to discovering the impact that mutational characteristics such as location, length of CT repeat, and parent of origin have on the symptomatic presentation of the disease, it is likely that relevant environmental factors will have to become part of the equation if genetic screening is to yield helpful and accurate predictions. This becomes

beliefs. Because of his opposition to abortion, Jêrome Lejeune, the French geneticist who discovered in 1958 that an extra chromosome causes Down's syndrome, has continued research into the biochemical causes of the condition. While "most geneticists regard [this] as a quixotic attempt to understand why a third 21-chromosome yields such debilitating results" (Kevles 1995, p. 288), it is motivated by Lejeune's goal to find some other "handle" by means of which to intervene in the treatment or prevention of the symptoms associated with Down's syndrome.

5.3.3 Causal Efficacy

Where variation in a trait in a population cannot be explained solely by genetic variation, causal efficacy provides a possible objective criterion for, nevertheless, designating the trait 'genetic.' Whether a "complex" qualitative trait like cardiovascular disease or a quantitative trait like I.Q. is involved, causal efficacy is gauged by efforts to quantify the relative contributions of various determining factors so as to identify the more "potent" cause or the cause with greater effect. Various approaches, both experimental and nonexperimental, are undertaken in the attempt to disentangle the genetic and nongenetic contributions to a particular trait and to estimate their relative importance. Experimental methods are best for establishing causal efficacy but obviously have limited applicability to humans. The goal is to quantify relations of functional dependence by measuring the effects of varying some properties while keeping others constant (Mackie 1980). In nonhuman organisms, molecular biologists exert technological control in the laboratory and quantitative geneticists exercise control over breeding. In human subjects, controlled clinical trials may be undertaken. Nonexperimental approaches, whether in human molecular genetics, epidemiology, or population genetics, measure variables of interest within a population and then attempt to estimate causal relations through determining the statistical significance of various correlations. For qualitative traits, where multiple factors are identified as statistically relevant to the outcome, one may be singled out as more important than the others because it confers the highest probability of sustaining a

---

even more critical as disease heterogeneity and the relative contribution of non-genetic factors increase.

given effect or because it is associated with the greatest proportion of cases in the population. Similarly, where estimates of heritability exceed the percentage of phenotypic variance attributed to the environment for a given quantitative trait, the trait may be labelled 'genetic.' In both experimental and nonexperimental settings and for qualitative and quantitative traits alike, whether or not various causes are additive in their effects is an important consideration in establishing causal relations and allocating causal responsibility.

Heritability coefficients are frequently appealed to in order to support claims for the causal efficacy of genes. For example, it is claimed that genes are more important than environment in determining behavioural characteristics because some recent twin studies have found 60% of phenotypic variation to be due to genetic variation and 40% to be due to environmental variation. Caution must be exercised, however, in making or accepting any generalizations about the causal effects of genes on a trait or a set of traits based on estimates of heritability. Besides the difficulties involved in obtaining accurate heritability measures in humans because of the inability to vary genetic and environmental backgrounds at will, heritability is a *local* statistical measure. The percentage of the total phenotypic variance that is due to genetic variance depends on the particular distribution of genotypes and environments in the population studied and it will fluctuate greatly between populations where there is significant gene-environment interaction. This is the basis of Lewontin's contention in his classic 1974 paper that it is impossible to infer causal relations from the analysis of variance. The only legitimate exception is where there is "perfect or nearly perfect additivity between genotypic and environmental effects so that the differences among genotypes are the same in all environments and the differences between environments are the same for all genotypes" (p. 408). Pointing to experimental evidence in nonhuman organisms, Lewontin argues that the effects of genetic and nongenetic causes are, more often than not, nonlinear. The question referred to as "Plomin's paradox" summarizes the responses of many behavioural geneticists: "If interactions are so ubiquitous in nature, why are they so difficult to find in behavioral research studies?" (Wachs 1991, p. 180). It may be that the focus on genetic main effects reflects the newness of the discipline and its lack, thus far, of the conceptual and methodological tools that will enable the effects of gene-environment interactions to be

detected (Plomin and Hershberger 1991, p. 29). However, there are also pragmatic considerations. While the lack of such tools may force researchers to concentrate on main (additive) effects and ignore interactions, behavioural geneticists pay little attention to main effects due to the environment. The failure to develop the necessary research protocols and statistical methods to measure interactions has a great deal to do with lack of interest, not only in interactionism, but in the environment as a whole.

Molecular geneticists have argued that the HGP does not ignore environmental influences on disease and behaviour; rather, after implicating as many relevant genetic factors as possible, it will be possible to delineate, and then to study, the causal role of the environment (Bodmer and McKie 1994). This assumes that we can understand the whole by partitioning its causal bits into those that are genetic and those that are environmental. But interactionism says that the whole is more (or less) than the sum of its parts. The question of whether the effects of individual causes are nonlinear and context-dependent or additive and context-independent is not just a feature unique to the statistical sampling of populations. If any attempt is to be made to quantify effects and to determine the relative contributions of different causal factors, the possible interactions of these factors must be considered in the experimental design. In the presence of significant gene-environment interaction, the magnitude and direction of the effects of manipulating the independent variable will depend on the specific values at which the dependent variables have been maintained constant. Any causal account of the development of traits will be partial, unless accompanied by a theoretical commitment to determining the full range of causal interactions.

5.4 Geneticization: Contexts and Choices

So far I have focused on the question "why single out genes as causes?" If we are to address the phenomenon of geneticization, that is, the increased frequency with which genetic explanations are offered for an expanding number of human conditions, it is necessary to ask a different, but related, question: "Why this increased singling out of genes as causes?" Many scientists explain geneticization by appeal to objective features

of the world. Watson and other molecular biologists believe that the development of genetic technologies has enabled the pivotal role of genes finally to be gleaned, for Mendelian and complex traits alike — recall the quotation concerning "the genetic foundation for cancer" referred to earlier. P. A. Baird (1990) argues that preventive efforts ought better be directed toward internal (genetic) factors in disease and away from external (environmental) ones on the basis of the increased relative importance of genes in morbidity as the incidence of infectious and nutritional diseases has dropped in developed countries. David Weatherall (1994) identifies widespread genetic susceptibilities to common diseases such as hypercholesteremia, diabetes, and obesity, borne of a hunter-gatherer genome unable to keep evolutionary pace with industrialization. In accepting that all traits are the product of gene-environment interaction and that the cause-condition distinction is always in part pragmatic, we are led to reject the epidemiological and evolutionary justifications of geneticization offered by Baird and Weatherall. "Handles" that promote successful intervention may be either genetic or environmental, not only in the common diseases referred to by Baird and Weatherall, but even in those considered to be caused by "single genes." Watson is right to focus on how the development of genetic technologies has changed our relationship with the world. This change, however, lies not in using newly acquired technological prowess to confirm the truth of long standing suspicions about the primacy of genes. I suggest, instead, that we understand geneticization in pragmatic terms: the increasing focus on genes as causes mirrors the increasing ability to manipulate DNA in the laboratory and in the clinic in furtherance of what are perceived to be desirable ends.

In effect, taking the pragmatic route bypasses interminable debates between realists and social constructivists. Explaining geneticization in terms of the "possibility and desirability" of manipulating genes does not deny, and in fact assumes, the materiality of DNA and the existence of an objective nexus of determining factors to which genes belong. It recognizes, no less, the paramount importance of the many contexts that shape explanatory aims in biology — the scientific, the clinical, the social, the economic, and the political — without the accompaniment of concomitant social constructivist claims that genes are invented fictions. Several theorists have explained geneticization in terms of changes in the social context. Edward Yoxen (1984) refers to the redefinition of

causes of disease that occurs as members of different medical specialties renegotiate their spheres of influence within institutional and professional structures. Hubbard (1990, with Wald 1993) cites wider social influences: the need to create and expand markets for the products of biotechnology; a preference for explanations of social inequities in terms of "innate" differences due to the conservative backlash to gains made by the civil rights and feminist movements; corporate and government disinclination to tackle the unhealthy environments associated with tobacco use, industrial pollutants, poverty, racism, etc.

It is not surprising that from laboratory contexts that treat genes as active causes and nongenetic factors as background conditions emerge theories that increasingly understand traits and diseases as 'genetic.' The result is to shift responsibility for disease from society to individual and to foster the belief that medical interventions would most successfully be directed at the level of the genome. But theory not only directs practice; it is directed by it. Causes, in this practical sense, are means to ends. Where there is a need for clinical intervention, the causal story must be appropriately simple — there must be some broken or missing part that can be replaced or substituted for by another. The contexts in which genes are chosen as the best "handles" among these parts are not just scientific and clinical, but economic and political. Geneticization finds a friendly home in a society less and less willing to commit resources to solving complex social problems. The perceived unwieldiness of items like poverty and pollution supports the molecular treatment of the environment as fixed and genes as active agents that can be localized and readily subdued by technological means. The search for quick and easy biotechnological fixes to complex problems is consonant with current economic priorities of governments motivated to reduce deficits by cutting spending. Genetically engineered solutions make private investors money; wars on drugs, poverty, environmental degradation just cost taxpayers money. An appreciation of the pragmatic dimensions of genetic explanations, and hence their contingency, not only provides good reason to be sceptical of what geneticization has to offer, but, by forcing attention to context, asks us to examine the aims, interests, and orientations that lie behind the choices that are being made. In this way, the debate between hereditarians and environmentalists is recast: the focus moves from questions concerning the veracity of different representations of reality to questions concerning preferences for certain kinds of interventions over others.

5.5 Summary

To argue, as I have, that traits are designated 'genetic' for pragmatic reasons is not to deny that genes are causally efficacious agents. We can speak sensibly about genetic causes and their effects, using either deterministic or probabilistic language, provided we recognize that we do so only relative to a particular set of background conditions, a specific population, and the present state of knowledge. What I do deny is that terms such as 'genetic trait,' 'genetic disease,' and 'genes for,' are objective, if we understand 'objective' to mean devoid of pragmatic content. I contend that how the cause-condition distinction is drawn, what population is selected, and which paths of research are followed, are choices that are influenced by the aims, interests, and orientations of those who make them. By appreciating the pragmatic dimensions of genetic explanations, we are forced to recognize their contingency and the need to interrogate the desires that shape the focus on genetic causes.

In van Fraassen's words: "scientific explanation is not (pure) science but an application of science. It is a use of science to satisfy certain of our desires; and these desires are quite specific in a specific context" (1980, p. 156). Opening the door onto context, we find that the focus on genetic causes satisfies many desires: scientific, technological, clinical, social, economic, and political. These coalesce in a single aim — that of control. Direct control over the hereditary material is now possible and the more cumbersome, unpredictable, and, for humans, unpalatable methods of control through selective breeding can be abandoned. On the horizon lies the potential for a plethora of desires to be satisfied: the eradication of human disease, control over human evolution, the genetic engineering of more appealing fruits and vegetables, the births of children with desirable qualities, etc. We cannot, however, look to a purely theoretical justification of such interventions in the designations of certain traits as 'genetic,' since it is these desires that lead us to focus on genes as causes in the first place.

Concluding Remarks: What is a "Normal" Genome and Does Anyone Have One?

The Human Genome Project is expected to culminate in the next seven years or so in the production of a single DNA sequence of over three billion nucleotide bases that is in some way supposed to represent the species. There are several ways in which this sequence might be conceived to be representative. As a composite of DNA sequences sampled from a small number of individuals, the reference sequence, although it represents no *actual* individual, could conceivably belong to *some* individual. In this way, it might be regarded as an "average" or "typical" human genome because, although it is likely that it never has existed and never will exist, it feasibly could exist. The sequence represents the species insofar as anything that is true of all human DNA sequences would be true of this one. If, as Walter Gilbert believes, there are genes that belong to all humans and only to humans, the nucleotides that comprise these genes would be contained in the reference sequence. If obtaining some such arbitrary DNA sequence is indeed the aim of the HGP, then my DNA sequence or your DNA sequence could equally be regarded as "the reality of our species" or as "the essential information that defines the type organism and hence the species." In fact, the genome of any individual whom we recognize to belong to *Homo sapiens* would suffice. Whether this person is healthy, diseased, or deformed should not matter.

This does not seem to be the HGP's intention, however. The goal to obtain the DNA sequence of a "type" organism is not to distinguish humans from nonhumans — we do this well at the gross phenotypic level already. Rather, the "type" organism defines a standard for intraspecific comparison. In this way, the reference sequence may represent either a statistical or a functional norm. A composite sequence compiled by stringing together sequences taken from a small number of individuals is an unlikely candidate to serve as a standard of statistical normality, however, given that it is likely that all humans who have ever lived, except in cases of monozygotic multiple births, have had unique DNA sequences. The composite DNA reference sequence that will be produced by the HGP will not even belong to an actual individual. Quite possibly, though, portions of the reference sequence that represent regions of the genome that are under functional constraint may be commonly shared by members of a population.

Whether this is indeed the case can only be established by the statistical sampling of populations. The status of a DNA reference sequence as a statistical norm would be strengthened by population studies to ensure that it represents at each locus and at each nucleotide position the most frequently occurring allele and nucleotide base, respectively. But even if molecular geneticists were prepared to carry out the extensive population research that this would require, the availability of a statistical standard of genetic normality is of questionable benefit. The reference sequence could represent *Homo sapiens* only by obliterating all genetic differences between populations and between individuals, whether these are adaptive or nonadaptive. The validity of a consensus sequence as a statistical representation of even a specific population is vulnerable to the same criticisms levelled at Quetelet's statistical conception of "*l'homme moyen*" or Galton's racial type. There is no guarantee that by joining together the most frequently occurring parts one obtains a whole that is itself common. The consensus sequence for a variant at a locus may not represent the most prevalent allele in the population. A genome that comprises the most frequently occurring alleles or nucleotides is even less likely to prevail in the population. In fact, it may not even exist.

It is difficult, in any case, to see what the purpose of a wholly statistical norm might be unless it is presumed that what is common also works well or is right and good in some other way. Certainly, it appears that the aim of producing a DNA reference sequence is to provide a genetic standard of normal functioning and health. As John Maddox writes in an article in *Nature* that makes "the case for the human genome": "The rapid and sure identification of genetic diseases by comparison between DNA sequences from the tissues of an affected person and some reference sequence in a databank is the most obvious benefit" (1991, p. 12). It is unlikely, however, that the individuals from whom the composite DNA reference sequence will be obtained carry not even a single allele that can be linked to disease or dysfunction in some genetic or environmental context. If the composite DNA reference sequence is to serve as a functional norm, it will need to undergo modification. One possible approach that, as we have seen, has been suggested with respect to the mtDNA reference sequence is to obtain a consensus sequence. Alternately, in the course of using the reference sequence in clinical studies

to identify disease mutations, polymorphisms associated with susceptibility to disease, and neutral variants, it will be possible to assess its validity as a standard of normal functioning on an ongoing basis and to replace any of its portions that are found to be associated with disease or dysfunction. Although, for both of these options, having a single standard of genetic normality seems to render all variation from the sequence suspect, it is not necessary that there be only one permissable "optimal" allele at each locus or nucleotide at each nucleotide position. Consensus sequences may include alternate nucleotides at positions along the sequence if these occur with adequate frequency in the relevant population. In the case of a composite sequence, data bases can be maintained on sequence variants which keep track of observed neutral polymorphisms.

If either of these approaches are to be successful, however, it is necessary that certain evolutionary assumptions be true. For a statistical norm (for example, a consensus sequence) to serve as a functional norm, there must be a constant relationship between an allele's or a nucleotide's frequency and its adaptive value. According to the classical account of the genetic structure of populations that is associated with H. J. Muller, there is a single optimally fit allele at each locus. Since it is assumed that stabilizing ("purifying") selection prevails and acts efficiently to eliminate even slightly inferior mutant alleles, the optimally fit allele will almost always be the allele that is found most frequently in the population. This is evidently not true where chance mechanisms in evolution predominate — for example, in areas of the genome that are of little functional importance and in small isolated populations. However, if we consider only functionally important regions of the genome and assume a large population, Motoo Kimura's neutral theory of molecular evolution supports the premise that frequently occurring alleles and nucleotides are fit although there is no reciprocal guarantee that infrequently occurring alleles and nucleotides are detrimental as they may be in mutation-drift, not mutation-selection, balance. If gene effects are almost always additive, as the classical and neutralist accounts purport, a composite of frequently occurring and therefore presumably fit alleles or nucleotides provides a suitable representation of a functionally normal genome. The second approach that uses clinical data to establish a suitable composite DNA reference sequence (with permissable functionally equivalent variants) similarly assumes that gene-gene and gene-environment interactions are overwhelmingly additive.

It is believed that an allele that is found to contribute to disease or dysfunction in one genetic and/or environmental context will do so in all genetic and/or environmental contexts. Certainly, as we have seen, the validity of these assumptions is in question. But, even according to Muller's and Kimura's own accounts, it is unlikely that a functionally normal genome exists in any actual individual due to the retention of deleterious mutant alleles where the effects of culture have led to a relaxation of natural selection or where effective population sizes have historically been relatively small. Today's human molecular geneticists assume that we are all more or less susceptible to various diseases and dysfunctions and that "virtually all human degenerative and infectious diseases are influenced by the genetic make-up of the individual" (Gottesman and Collins 1994, p. 591).

Hence, whether we take the "ultimate" DNA reference sequence to be the composite DNA sequence in which the HGP culminates or a subsequent modification that is based on population studies of the frequency and/or functionality of specific alleles or nucleotides, it is doubtful that the entire sequence will be found to be instantiated in any actual individual. Whether the reference sequence is conceived as an arbitrary human genome, a statistical norm, or a functional norm, it *could* feasibly exist, but likely does not. Although Cournot attacked Quetelet's concept of *l'homme moyen* and Dobzhansky attacked Muller's "all-normal man" on these grounds, this seems inadequate reason to dismiss entirely the validity of the DNA reference sequence's representational status since many representations are, after all, idealizations. A statistical norm that consists of the most frequently occurring allele at each locus or nucleotide at each nucleotide position, or a functional norm that excludes all alleles that are known to be associated with disease or dysfunction, might still serve as a valid standard for comparison even if there is no actual genome that corresponds with either *in toto*. The concept of the "normal" genome as ideal genome receives theoretical content from two sources — evolutionary and clinical.

From the evolutionary perspective, the "normal" genome is the unblemished "original." In Chapter Three, I argued that Muller's classical and Kimura's neutralist accounts of the genetic structure of populations support a conception of the "normal" genome as the "original" genome. Muller believed that the optima for physical traits are

to be found in our hunter-gatherer past and it seems that Kimura goes back even further than this. Mutations are not simply changes to a preexisting genetic structure. Rather, since the "original" genome is considered be optimally adaptive, mutations constitute damage. However, there is no "normal" genome as "original" genome without several accompanying assumptions. For both Muller and Kimura, natural selection is mostly a negative stabilizing force that preserves phenotypic form and function by eliminating deviates. Population size is large, environments are relatively constant over evolutionary time, and homeostatic physiological mechanisms buffer organisms from short-term fluctuations in the environment. Alleles have absolute selective values: they are good, bad, or indifferent in perpetuity. For Muller, hunter-gatherer environments provided conditions that were ideal for natural selection to shape and to preserve a species-adaptive norm. Environmental changes associated with the cultural progress that has occurred over the past few millenia mean that some "deleterious" alleles are no longer detrimental to fitness and are consequently retained in the population. Similarly, for Kimura, small effective population sizes mean that "deleterious" alleles accumulate in frequency because chance prevails over selection.

From the clinical perspective, the "normal" genome lacks all mutations that are associated with conditions that are disvalued — whether, as I argued in Chapter Four, for biological or sociocultural reasons. According to the etiological account of biological functions, the "normal" environment is the historical one in which a particular allele was selected whereas, according to the dispositional account, the "normal" environment is the present one. In the event of environmental change, the same allele may be considered to be normal from one perspective and abnormal from another. Civilization may be viewed a threat to the biological health of the species or evolution by natural selection may be regarded as too slow of an accommodation of cultural progress. But, in either case, reproductive success is an inadequate criterion for upon which to base judgements of health and disease because humans are cultural, as well as biological, beings who experience environments that are both social and physical. Any values — aesthetic, moral, social, or cultural, as well as biological — that attach to clinical judgements of health and disease at the levels of individuals in their environments are incorporated into judgements of normal and abnormal gene function. The "normal" genome is not the

unblemished "original," as it is understood from the evolutionary perspectives of Muller and Kimura, but the envisioned future creation.

Hence, either from an evolutionary or a clinical perspective, the concept of the "normal" genome as the ideal genome can be given theoretical content in support of the reference sequence's use in directing and sanctioning genetic interventions. The "normal" genome is a *desideratum.* Although no actual individual may have a "normal" genome, today, with the availability of cut-and-paste recombinant DNA technologies, it has become possible to close the gap between idea and reality. The "ultimate" map, the DNA reference sequence, potentially provides a set of instructions for the technological modification of existing genomes. As the 1988 U.S. Congress' Office of Technology Assessment report on the HGP notes, "new technologies for identifying traits and altering genes make it possible for eugenic goals to be achieved through technological as opposed to social control" (in Keller 1992, p. 295). The report refers favourably to a new "eugenics of normalcy" that ensures the "paramount right" of each individual to be born with "at least a modicum of normal genes" (ibid.). This "modicum of normal genes" can be understood in either of the two senses of the "normal" genome as the ideal genome. As we saw in Chapter Four, some bioethicists argue that germ-line manipulation has ethical warrant if it aims only to "restore" an "'original' healthy genetic topology" that has been "disturbed by genetic mutation." Appeal is made also to the possible fashioning of entirely new genes and traits. Recall Robert Sinsheimer's 1969 forecast, just as the technological revolution in molecular biology was beginning, of the promise of a "new eugenics" that would, in theory, permit all individuals to be converted to "the highest technological level."

A potentially dangerous combination of factors presents: the gap between the idea of the "normal" genome and reality, the intersection of multiple senses of 'normal' in the concept of the "normal" genome, the eugenic aims associated with the HGP, and the power of the new genetic technologies. This calls to mind a passage from Ian Hacking's *The Taming of Chance*:

> On the one hand there is the thought that the normal is what is right, so that talk of the normal is a splendid way of preserving or returning to the status quo.... On the other hand is the idea that the normal is only average, and so is something to be improved upon.... The normal stands indifferently for what is typical, the

> unenthusiastic objective average, but it also stands for what has been, good health, and for what shall be, our chosen destiny. That is why the benign and sterile-sounding word 'normal' has become one of the most powerful ideological tools of the twentieth century. (1990, pp. 168-169)[1]

The HGP's aim to produce a DNA reference sequence *does* receive theoretical support from the concept of the "normal" genome as the ideal genome. However, the validity of the assumptions that support Muller's and Kimura's accounts of the genetic structure of populations and the evolutionary concept of the "normal" genome as the "original" genome is in doubt. Even if these assumptions were true, it would be a near-impossible task to pick out the "original" allele from among all the variants observed to segregate at a given locus in order to establish a reference sequence. Muller's belief that the environment has deteriorated as a result of civilization and Kimura's view that chance mechanisms have prevailed in evolution due to relatively small population sizes alike entail that the allele that occurs most frequently in human populations is not necessarily the most fit and/or the oldest. Instead, we are dependent on determinations of how alleles function in present environments. As we have seen, clinical judgements of health and disease and the identification of mutant or abnormal genes are prior to, and constitutive of, judgements of normal gene function. The designation of a particular allele as normal or abnormal depends on what we take to be a normal or abnormal phenotype and a normal or abnormal environment. Since human environments are social as well as physical, such judgements incorporate both biological and nonbiological values and will vary from one culture to another.

Hence, although the concept of the "normal" genome as the ideal genome has theoretical content, it is virtually impossible to establish a definitive human DNA reference sequence. This impossibility supports the substitution of an engineering norm for an empirical (or scientific) norm. Nothing circumvents proceeding from interventions that repair malfunctioning machines to ones that build "better" functioning ones. Any judgement of what constitutes a "better" genotype is entirely dependent on what we take to be desirable phenotypes and environments and these desires will be shaped by social

---

[1] Keller (1992) has already drawn this connection between the OTA report, Sinsheimer's vision, and Hacking's account.

and cultural, as well as biological, values. The result is a concept of a "normal" genome that is always able to accommodate changing human desires. As Sinsheimer predicted, the technological capacity to close the gap between idea and reality, "to bring everyone to the highest technological level," now exists. The elasticity of the concept of a "normal" genome combined with the forces of market economies driven by "for-profit" health care, biotechnology investments, and consumer demands for "better living" and "better babies" may well guarantee the maintenance of such a gap.

A DNA reference sequence need not be considered to be authoritative. It is possible to regard departures from the sequence as only *potential* candidates for classification as harmful deviates, the judgement of which requires additional functional information. While this nonauthoritative use of the reference sequence suffices where genetic screening is carried out for diagnostic purposes in existing individuals, practical decisions based on DNA sequence data that concern not-yet-existing individuals for whom this additional functional information is unavailable — for example, in germ-line manipulation, IVF embryo selection, or selective abortion — must necessarily treat the reference sequence as authoritative. And, yet, there is real difficulty in establishing any such definitive standard of genetic normality. The challenge that faces us is to develop new and different ways of understanding human genetic variation. Although the HGP has been criticized from the start for ignoring genetic variation, only recently has any attempt been made to begin to collect data on variation. However, these attempts lie firmly within the biomedical framework that understands human variation as deviation or deviance. A "Mutation Database Initiative" was started in 1994 and became part of HUGO in 1997. Locus-specific mutations are named according to the nucleotide "change" that has occurred; in other words, mutations are regarded as departures from *some* "original" normal allele (Cotton *et al.* 1998). In another initiative, Francis Collins suggests that a genetic map composed of SNP markers obtained by sampling 100-500 African-Americans, Asian-Americans, European-Americans, and Native Americans will not only help to locate new genes but will identify variant forms of known genes and maybe even help the HGDP to get started. Although the HGDP seeks to uncover knowledge of human evolutionary history, its proponents have also presented it as a panacea for the lack of attention paid by the HGP to the study of human genetic variation.

As biological anthropologist Kenneth M. Weiss (1996) writes, the HGDP will contribute to "our understanding of *normal* human variation and its origins, a subject too often omitted in biomedical research" (p. 293). But, as I argued in Chapter Three, a populations approach is no guarantee that genetic variation will be conceived as difference rather than as deviation. In addition, it is no easier to draw a line between normal and abnormal genetic variation than it is to draw one between normal and abnormal genes. Multifactorial diseases and complex behavioural traits are likely to yield to "molecular dissection." There is no reason why what is considered to be "normal" variation today will not become parcelled into "acceptable" and "unacceptable" components tomorrow.

The April 1998 cover of *Life* magazine catches the eye of the newstand browser. A brightly coloured double helix is accompanied by the headline's question: "WERE YOU BORN THAT WAY? Personality, temperament, even life choices. New studies show it's mostly in your genes." Portrayed in the photographs that *Life* is famous for are a shy four-year-old girl and her once-shy mother, a thrill-seeking TV stuntman, a mother and daughter who are both obese, an active five-year-old boy referred to in the caption as "testosterone-driven" by his mother, a gay couple, and a male smoker who is also a recovering heroin-addict and alcoholic. If people believe that "solutions" to their "problems" are technological and pharmaceutical, the trend to medicalize human characteristics that were previously considered to be moral or social — alcoholism and drug abuse, for example — will continue. Genetic variation can be alternately understood as difference, deviation, or deviance, and there is no incontrovertible principled distinction that might rule between these.

References

Abelson, Philip H. 1996. "Pharmaceuticals Based on Biotechnology." *Science* **273** (9 August): 719.

Adams, Mark B. 1968. "The Founding of Population Genetics: Contributions of the Chetverikov School, 1924-1934." *Journal of the History of Biology* **1**: 23-39.

Adams, Mark D., et al. 1995. "Initial Assessment of Human Gene Diversity and Expression Patterns Based on 83 Million Nucleotides of cDNA Sequence." *Nature* **377** supp. (28 Sept): 3-17.

Agich, George J. 1983. "Disease and Value: A Rejection of the Value-Neutrality Thesis." *Theoretical Medicine* **4**: 27-41.

Alberts, Bruce, et al. 1998. *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell.* New York: Garland Publishing.

Allen, Garland E. 1978. *Thomas Hunt Morgan: The Man and His Science.* Princeton, N.J.: Princeton University Press.

Anderson, Christopher. 1993. "Genome Project Goes Commercial." *Science* **259** (15 January): 300-302.

Anderson, S., et al. 1981. "Sequence and Organization of the Human Mitochondrial Genome." *Nature* **290**: 457-465.

Annas, George J. 1989. "Who's Afraid of the Human Genome?" *Hastings Center Report* (July/August): 19-21.

Aquadro, Charles F., and Barry D. Greenberg. 1983. "Human Mitochondrial DNA Variation and Evolution: Analysis of Nucleotide Sequences from Seven Individuals." *Genetics* **103**: 287-312.

Avery, Oswald T., Colin M. MacLeod, and Maclyn McCarty. 1944. "Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III." *Journal of Experimental Medicine* **79** (1 February): 137-158.

Baird, P. A. 1990. "Genetics and Health Care: A Paradigm Shift." *Perspectives in Biology and Medicine* **33**: 203-213.

Beatty, John. 1987. "Weighing the Risks: Stalemate in the Classical/Balance Controversy." *Journal of the History of Biology* **20**: 289-319.

Bentley, David R. 1996. "Genomic Sequence Information Should Be Released Immediately and Freely in the Public Domain." *Science* **274** (25 October): 533-536.

Berg, Paul. 1993. "Co-Chairman's Remarks: Reverse Genetic: Directed Modification of DNA for Functional Analysis." *Gene* **135**: 261-264.

Berkowitz, Ari. 1996. "Our genes, ourselves?" *BioScience* **46**: 42-51.

Bernard, Claude. [1927] 1957. *An Introduction to the Study of Experimental Medicine.* Translated by Henry Copley Greene. Reprint, with a forward by I. Bernard Cohen, New York: Dover Publications. Originally published as *Introduction à l'Etude de la Médecine Experimentale* (Paris: J. B. Baillière, 1865).

Bigelow, John, and Robert Pargetter. 1987. "Functions." *The Journal of Philosophy* **84**: 181-196.

Bodmer, Walter. 1986. "Human Genetics: The Molecular Challenge." *Cold Spring Harbor Symposia on Quantitative Biology* **51**: 1-13.

Bodmer, Walter, and Robin McKie. 1994. *The Book of Man: The Quest to Discover Our Genetic Heritage.* Toronto: Viking Press.

Boorse, Christopher. 1977. "Health as a Theoretical Concept." *Philosophy of Science* **44**: 542-573.

Boorse, Christopher. 1984. "Wright on Functions." In *Conceptual Issues in Evolutionary Biology: An Anthology*, edited by Elliot Sober, 369-385. Cambridge, Mass.: The MIT Press. First published in *Philosophical Review* **85** (1976): 70-86.

Bowler, Peter J. 1989. *The Mendelian Revolution: The Emergence of Hereditarian Concepts in Modern Science and Society.* London: The Athlone Press.

Bunzl, Martin. 1980. "Comment on 'Health as a Theoretical Concept.'" *Philosophy of Science* **47**: 116-118.

Butler, Declan. 1995. "Genetic Diversity Proposal Fails to Impress International Ethics Panel." *Nature* **377** (5 October): 373.

Butler, Judith. 1990. *Gender Trouble: Feminism and the Subversion of Identity.* New York: Routledge.

Canguilhem, Georges. [1978] 1989. *The Normal and the Pathological.* Translated by Carolyn R. Fawcett in collaboration with Robert S. Cohen. Reprint, New York: Zone Books. Originally published as *Le Normal et le Pathologique*, being in part the author's thesis, 1943 (Paris: Presses Universitaires de France, 1966).

Cann, Rebecca L., Mark Stoneking, and Allan C. Wilson. 1987. "Mitochondrial DNA and Human Evolution." *Nature* **325**: 31-36.

Cantor, Charles R. 1990. "Orchestrating the Human Genome Project." *Science* **248** (6 April): 49-51.

Carlson, Elof Axel, ed. 1973. Introduction to *Man's Future Birthright: Essays on Science and Humanity*, by H. J. Muller. Albany: State University of New York Press.

Casalino, Larry. 1991. "Decoding the Human Genome Project: An Interview with Evelyn Fox Keller." *Socialist Review* **21** (2): 111-128.

Caskey, C. Thomas. 1992. "DNA-Based Medicine: Prevention and Therapy." In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, edited by Daniel J. Kevles and Leroy Hood, 112-135. Cambridge: Harvard University Press.

Chargaff, Edwin. 1950. "Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymic Degradation." *Experientia* **6**: 201-209.

Chumakov, Ilya M., et al. 1995. "A YAC Contig Map of the Human Genome." *Nature* **377** supp. (28 September): 175-183.

Cohen, I. Bernard. 1987. "Scientific Revolutions, Revolutions in Science, and a Probabilistic Revolution, 1800-1930." In *The Probabilistic Revolution*. Vol. 1, *Ideas in History*, edited by Lorenz Krüger, Lorraine J. Daston, and Michael Heidelberger, 23-44. Cambridge: The MIT Press.

Cohen, Jon. 1997. "How Many Genes Are There?" *Science* **275**: 769.

Coleman, William. 1987. "Experimental Physiology and Statistical Inference: The Therapeutic Trial in Nineteenth-Century Germany." In *The Probabilistic Revolution*. Vol. 2, *Ideas in Science*, edited by Lorenz Krüger, Gerd Gigerenzer, and Mary S. Morgan, 201-226. Cambridge: The MIT Press.

Collingwood, R. G. 1938. "On the So-Called Idea of Causation." *Proceedings of the Aristotelian Society*: 85-112.

Collins, Andrew R. 1996. "The Nature and Significance of Mutations." In *Molecular and Cellular Genetics*, edited by E. Edward Bittar and Neville Bittar, 255-270. Greenwich, CT: JAI Press.

Collins, Francis S. 1995. "Positional Cloning Moves from Perditional to Traditional." *Nature Genetics* **9**: 347-350.

Collins, Francis, and David Galas. 1993. "A New Five-Year Plan for the U.S. Human Genome Project." *Science* **262** (1 October): 43-46.

Collins, Francis S., Mark S. Guyer, and Aravinda Chakravarti. 1997. "Variations on a Theme: Cataloging Human DNA Sequence Variation." *Science* **278** (28 November): 1580-1581.

Cook-Deegan, Robert. 1994. *The Gene Wars: Science, Politics, and the Human Genome*. New York: W. W. Norton & Co.

Cooper, D. N., and M. Krawczak. 1993. *Human Gene Mutation*. Oxford: BIOS Scientific Publishers.

Cordeiro, A. R. and Theodosius Dobzhansky. 1954. "Combining Ability of Certain Chromosomes in Drosophila Willistoni and Invalidation of the 'Wild-Type' Concept." *The American Naturalist* **88**: 75-86.

Côrte-Real, H. B. S. M., et al. 1996. "Genetic Diversity in the Iberian Peninsula Determined from Mitochondrial Sequence Analysis. *Annals of Human Genetics* **60**: 331-350.

Cotton, R. G. H., V. McKusick, and C. R. Scriver. 1998. "The HUGO Mutation Database Initiative." *Science* **279** (2 January): 10-11.

Crick, Francis. 1993. "Looking Backwards: A Birthday Card for the Double Helix." *Gene* **135**: 15-18.

Crow, James F. 1972. "The Dilemma of Nearly Neutral Mutations. How Important Are They for Evolution and Human Welfare?" *Journal of Heredity* **63**: 306-316.

Crow, James F. 1981. "The Neutralist-Selectionist Controversy: An Overview." In *Population and Biological Aspects of Human Mutation*, edited by Ernest B. Hook and Ian H. Porter, 3-14. New York: Academic Press.

Crow, James F. 1985. "The Neutrality-Selection Controversy in the History of Evolution and Population Genetics." In *Population Genetics and Molecular Evolution: Papers Marking the Sixtieth Birthday of Motoo Kimura*, edited by Tomoko Ohta and Kenichi Aoki, 1-18. Tokyo: Japan Scientific Societies Press.

Crow, James F. 1987. "Muller, Dobzhansky, and Overdominance." *Journal of the History of Biology* **20**: 351-380.

Cummins, Robert. 1984. "Functional Analysis." In *Conceptual Issues in Evolutionary Biology: An Anthology*, edited by Elliot Sober, 386-407. Cambridge, Mass.: The MIT Press. First published in *Journal of Philosophy* **72** (1975): 741-764.

Daston, Lorraine J. 1987. "Rational Individuals versus Laws of Society: From Probability to Statistics." In *The Probabilistic Revolution*. Vol. 1, *Ideas in History*, edited by Lorenz

Krüger, Lorraine J. Daston, and Michael Heidelberger, 295-304. Cambridge: The MIT Press.

Dawkins, Richard. 1982. *The Extended Phenotype: The Long Reach of the Gene.* Oxford: Oxford University Press.

Diamond, Solomon. 1969. Introduction to *A Treatise on Man and the Development of his Faculties*, by Lambert A. J. Quetelet. Gainsville: Scholars' Facsimiles & Reprints.

Dib, C., et al. 1996. "A Comprehensive Genetic Map of the Human Genome based on 5,264 Microsatellites." *Nature* **380** (14 March): 152-154.

Dietrich, Michael R. 1994. "The Origins of the Neutral Theory of Molecular Evolution." *Journal of the History of Biology* **27**: 21-59.

Dobzhansky, Theodosius. 1955a. "A Review of Some Fundamental Concepts and Problems of Population Genetics." *Cold Spring Harbor Symposia of Quantitative Biology* **20**: 1-15.

Dobzhansky, Theodosius. 1955b. *Evolution, Genetics, and Man.* New York: John Wiley & Sons.

Dobzhansky, Theodosius. 1962. *Mankind Evolving: The Evolution of the Human Species.* New Haven: Yale University Press.

Dobzhansky, Theodosius. 1973. "Nothing in Biology Makes Sense Except in the Light of Evolution." *American Biology Teacher* **35**: 125-129.

Dobzhansky, Theodosius. 1976. "Organismic and Molecular Aspects of Species Formation." In *Molecular Evolution*, edited by Francisco J. Ayala, 95-105. Sunderland, Mass.: Sinauer Associates.

Ducor, Philippe. 1997. "Are patents and research compatible?" *Nature* **387** (1 May): 13-14.

Dulbecco, Renato. 1996. "A Turning Point in Cancer Research: Sequencing the Human Genome." *Science* **231** (7 March): 1055-1056.

Dulbecco, Renato. 1993. "Co-Chairman's Remarks: The Human Genome Project and Cancer." *Gene* **135**: 259-260.

Duster, Troy. 1990. *Backdoor to Eugenics.* New York: Routledge.

"Editorial." 1994. *Nature Genetics* **7** (2 special issue).

Edlin, Gordon J. 1987. "Inappropriate Use of Genetic Terminology in Medical Research: A Public Health Issue." *Perspectives in Biology and Medicine* **31**: 47-56

Engelhardt, H. Tristram, Jr. 1981. "The Concepts of Health and Disease." In *Concepts of Health and Disease: Interdisciplinary Perspectives*, edited by Arthur L. Caplan, H. Tristram Engelhardt, Jr., and James J. McCartney. London: Addison-Wesley Publishing Co. First published in H. Tristram Engelhardt, Jr. and Stuart F. Spicker, *Evaluation and Explanation in the Biomedical Sciences* (Dordrecht: D. Reidel, 1975).

Fausto-Sterling, Anne. 1992. *Myths of Gender: Biological Theories About Women and Men*. 2nd ed. New York: Basic Books.

Forest, Maguelone G., ed. 1989. *Androgens in Childhood: Biological, Physiological, Clinical and Therapeutic Aspects*. Pediatric and Adolescent Endocrinology, vol. 19. Basel: S. Karger.

Foucault, Michel. [1978] 1989. Introduction to *The Normal and the Pathological*, by Georges Canguilhem. New York: Zone Books.

Foucault, Michel. 1990. *The History of Sexuality*. Vol. 1, *An Introduction*. Translated by Robert Hurley. New York: Vintage Books.

Freese, Ernst. 1963. "Molecular Mechanism of Mutation." In *Molecular Genetics*, part 1, edited by J. Herbert Taylor. New York: Academic Press.

Galton, Francis. [1869] 1979. *Hereditary Genius: An Enquiry Into Its Laws and Consequences*. Reprint, with a forward by H. J. Eysenck and the author's preface from the 1892 ed., London: Julian Friedmann Publishers.

Galton, Francis. 1877. "Typical Laws of Heredity." *Nature* (12 April): 512-513.

Galton, Francis. 1907. *Probability, the Foundations of Eugenics*. Oxford: Clarendon Press.

Gebhard, W., et al. 1982. "A New Family of Interspersed Repetitive DNA Sequences in the Mouse Genome." *Journal of Molecular Biology* **157**: 453-472.

Gifford, Fred. 1990. "Genetic Traits." *Biology and Philosophy* **5**: 327-347.

Gilbert, Walter. 1992. "A Vision of the Grail." In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, edited by Daniel J. Kevles and Leroy Hood, 83-97. Cambridge, Mass.: Harvard University Press.

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

Ginther, Charles, Laurie Issel-Tarver, and Mary-Claire King. 1992. "Identifying Individuals by Sequencing Mitochondrial DNA from Teeth." *Nature Genetics* **2**: 135-138.

Gottesman, Michael M., and Francis S. Collins. 1994. "The Role of the Human Genome Project in Disease Prevention." *Preventive Medicine*. **23**: 591-594.

Gould, Stephen Jay. 1983. "The Hardening of the Modern Synthesis." In *Dimensions of Darwinism: Themes and Counterthemes in Twentieth-Century Evolutionary Theory*, edited by In Marjorie Grene, 71-93. Cambridge, Cambridge University Press.

Griesemer, James R. 1994. "Tools for Talking: Human Nature, Weismannism, and the Interpretation of Genetic Information." In *Are Genes Us?: The Social Consequences of the New Genetics*, edited by Carl F. Cranor. New Brunswick, N.J.: Rutgers University Press.

Guyer, Ruth Levy, and Daniel E. Koshland. 1989. "The Molecule of the Year." *Science* **246** (22 December): 1543-1544.

Hacking, Ian. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.

Haldane, J. B. S. 1937. "The Effect of Variation on Fitness." *American Naturalist* **71**: 337-349.

Halling, S. M., and N. Kleckner. 1982. "A Symmetrical 6-Base-Pair Target Site Sequence Determines Transposon Tn10 Insertion Specificity." *Cell* **28**: 155-164.

Hardman, J. A., et al. 1984. "Primary Structure of the Human Renin Gene." *DNA* **3**: 457-468.

Harris, Harry. 1966. "Enzyme polymorphism in man." *Proceedings of the Royal Society of London, ser. B* **164**: 298-310.

Hasegawa, Masami and Satoshi Horai. 1991. "Time of the Deepest Root for Polymorphism in Human Mitochondrial DNA." *Journal of Molecular Evolution* **32**: 37-42.

Heerdt, Barbara G., et al. 1994. "Polymorphisms, but Lack of Mutations or Instability, in the Promotor Region of the Mitochondrial Genome in Human Colonic Tumors." *Cancer Research* **54**: 3912-3915.

Hershey, A. D., and Martha Chase. 1952. "Independent Functions of Viral Protein and Nucleic Acid in Growth of Bacteriophage." *Journal of General Physiology* **36**: 39-56.

Hesslow, Germund. 1983. "Explaining and Weighting Causes." *Theoria* **49**: 87-111.

Hesslow, Germund. 1984. "What is a Genetic Disease?: On the Relative Importance of Causes." In *Health, Disease, and Causal Explanations in Medicine*, edited by Lennart Nordenfelt and B. Ingemar B. Lindhal, 183-193. Dordecht: D. Reidel.

Hesslow, Germund. 1993. "Do We Need a Concept of Disease?" *Theoretical Medicine* **14**: 1-14.

Ho, Lap, and Mulchand S. Patel. 1990. "Cloning and cDNA Sequence of the β-subunit Component of Human Pyruvate Dehydrogenase Complex." *Gene* **86**: 297-302.

Ho, Lap, et al. 1989. "Characterization of cDNAs Encoding Human Pyruvate Dehydrogenase α-subunit." *Proceedings of the National Academy of Sciences USA* **86**: 5330-5334.

Holland, Mitchell M., et al. 1993. "Mitochondrial DNA Sequence Analysis of Human Skeletal Remains: Identification of Remains from The Vietnam War." *Journal of Forensic Sciences* **38**: 542-553.

Hood Leroy. 1992. "Biology and Medicine in the Twenty-First Century." In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, edited by Daniel J. Kevles and Leroy Hood, 136-163. Cambridge, Mass.: Harvard University Press.

Horai, Satoshi, and Kenji Hayasaka. 1990. "Intraspecific Nucleotide Sequence Differences in the Major Noncoding Region of Human Mitochondrial DNA." *American Journal of Human Genetics* **46**: 828-842.

Horai, Satoshi, et al. 1992. "Man's Place in Hominoidea Revealed by Mitochondrial DNA Genealogy." *Journal of Molecular Evolution* **35**: 32-43.

Horai, Satoshi, et al. 1995. "Recent African Origin of Modern Humans Revealed by Complete Sequences of Hominoid Mitochondrial DNAs." *Proceedings of the National Academy of Sciences USA* **92**: 532-536.

Horan, Barbara L. 1989. "Functional Explanations in Sociobiology." *Biology and Philosophy* **4**: 131-158.

Howell, Neil, et al. 1992. "The Sequence of Human mtDNA: The Question of Errors versus Polymorphisms." *American Journal of Human Genetics* **50**: 1333-1337.

Hubbard, Ruth. 1990. *The Politics of Women's Biology*. New Brunswick, NJ: Rutgers University Press.

Hubbard, Ruth, and Elijah Wald. 1993. *Exploding the Gene Myth: How Genetic Information is Produced and Manipulated by Scientists, Physicians, Employers, Insurance Companies, Educators, and Law Enforcers*. Boston: Beacon Press.

Hubby, J. L. and R. C. Lewontin. 1966. "A Molecular Approach to the Study of Genic Heterozygosity in Natural Populations. I. The Number of Alleles at Different Loci in Drosophila Pseudoobscura." *Genetics* **54**: 577-594.

Hull, David L. 1994. "Species, Races, and Genders: Differences Are Not Deviations." In *Genes and Human Self-Knowledge: Historical and Philosophical Reflections on Modern Genetics*, edited by Robert F. Weir, Susan C. Lawrence, and Evan Fales, 207-219. Iowa City: University of Iowa Press.

Jaenisch, Rudolf. 1988. "Transgenic Animals." *Science* **240**: 1468-1473.

James, Vivian H. T., ed. 1992. *The Adrenal Gland.* 2nd ed. New York: Raven Press.

Jordan, Elke, and Francis S. Collins. 1996. "A march of genetic maps." *Nature* **380** (14 March): 111-112.

Josso, Natalie, ed. 1981. *The Intersex Child.* Pediatric and Adolescent Endocrinology, vol. 8. Basel: S. Karger.

Judson, Horace Freeland. 1992. "A History of the Science and Technology Behind Gene Mapping and Sequencing." In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, edited by Daniel J. Kevles and Leroy Hood, 36-80. Cambridge: Harvard University Press.

Judson, Horace Freeland. 1996. *The Eighth Day of Creation: Makers of the Revolution in Biology.* Expanded ed. New York: Cold Spring Harbor Laboratory Press.

Jun, Albert S., Michael D. Brown, and Douglas C. Wallace. 1994. "A Mitochondrial DNA Mutation at Nucleotide Pair 14459 of the NADH Dehydrogenase Subunit 6 Gene Associated with Maternally Inherited Leber Hereditary Optic Neuropathy and Dystonia. *Proceedings of the National Academy of Sciences USA* **91**: 6206-6210.

Kahn, Patricia. 1994. "Genetic Diversity Project Tries Again." *Science* **266** (4 November): 720-722.

Kahn, Patricia. 1996. "The Right Mix of Form and Function." *Science* **273** (2 August): 570-571.

Kaplan, Abraham. 1965. "Noncausal Explanation." In *Cause and Effect: The Hayden Colloquium on Scientific Method and Concept*, edited by Daniel Lerner, 145-155. New York: The Free Press.

Kaiser, Jocelyn, ed. 1997. "ScienceScope." *Science* **277** (18 July): 303.

Kearsey, Michael J., and Harpal S. Pooni. 1996. *The Genetical Analysis of Quantitative Traits.* London: Chapman and Hill.

Keller, Evelyn Fox. 1990. "Physics and the Emergence of Molecular Biology: A History of Cognitive and Political Synergy." *Journal of the History of Biology* **23**: 389-409.

Keller, Evelyn Fox. 1992. "Nature and Nurture." In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, edited by Daniel J. Kevles and Leroy Hood, 281-299. Cambridge, Mass.: Harvard University Press.

Keller, Evelyn Fox. 1994. "Master Molecules." In *Are Genes Us?: The Social Consequences of the New Genetics*, edited by Carl F. Cranor, 89-98. New Brunswick, N.J.: Rutgers University Press.

Keller, Evelyn Fox. 1995. *Refiguring Life: Metaphors of Twentieth-Century Biology*. New York: Columbia University Press.

Kevles, Daniel J. 1995. *In the Name of Eugenics: Genetics and the Uses of Human Heredity*. 2nd ed. Cambridge, Mass.: Harvard University Press.

Kimura, Motoo. 1968. "Evolutionary Rate at the Molecular Level." *Nature* **217** (17 February): 624-626.

Kimura, Motoo. 1979a. "Model of Effectively Neutral Mutations in which Selective Constraint is Incorporated." *Proceedings of the National Academy of Sciences USA* **76**: 3440-3444.

Kimura, Motoo. 1979b. "The Neutral Theory of Molecular Evolution." *Scientific American* **241** (5): 98-126.

Kimura, Motoo. 1982. "The Neutral Theory as a Basis for Understanding the Mechanism of Evolution and Variation at the Molecular Level." In *Molecular Evolution, Protein Polymorphism and the Neutral Theory*, edited by Motoo Kimura, 3-56. Tokyo: Japan Scientific Societies Press.

Kimura, Motoo. 1983a. "Rare Variant Alleles in the Light of the Neutral Theory." *Molecular Biology and Evolution* **1**: 84-93.

Kimura, Motoo. 1983b. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kimura, Motoo. 1986. "DNA and the Neutral Theory." *Philosophical Transactions of the Royal Society of London, ser. B* **312**: 343-354.

Kimura, Motoo. 1991. "Some Recent Data Supporting the Neutral Theory." In *New Aspects of the Genetics of Molecular Evolution*, edited by M. Kimura and N. Takahata, 3-14. Tokyo: Japan Scientific Society Press.

King, J. L. 1967. "Continuously Distributed Factors Affecting Fitness." *Genetics* **55**: 483-492.

King, J. L. and T. H. Jukes. 1969. "Non-Darwinian Evolution: Random Fixation of Selectively Neutral Mutations. *Science* **164**: 788-798.

Kitcher, Philip. 1996. *The Lives to Come: The Genetic Revolution and Human Possibilities*. New York: Simon & Schuster.

Koenig, Robert. 1996. "Giant Merger Creates Biotech Power." *Science* **271** (15 March): 1490.

Koenig, Robert. 1997. "Watson Urges 'Put Hitler Behind Us.'" *Science* **276** (9 May): 892.

Kogelnik, Andreas M. 1996. "MITOMAP: A Human Mitochondrial Genome Database." *Nucleic Acids Research* **24**: 177-179.

Kohler, Robert E. 1994. *Lords of the Fly:* Drosophila *Genetics and the Experimental Life*. Chicago: University of Chicago Press.

Kozak, M. 1986. "Point Mutations Define a Sequence Flanking the AUG Initiator Codon that Modulates Translation by Eukaryotic Ribosomes." *Cell* **44**: 283-292.

Krüger, Lorenz. 1987. "The Slow Rise of Probabilism: Philosophical Arguments in the Nineteenth Century." In *The Probabilistic Revolution*. Vol. 1, *Ideas in History*, edited by Lorenz Krüger, Lorraine J. Daston, and Michael Heidelberger, 59-89. Cambridge, Mass.: The MIT Press.

Lander, Eric S., and David Botstein. 1986. "Mapping Complex Traits in Humans: New Methods Using a Complete RFLP Linkage Map." *Cold Spring Harbor Symposia on Quantitative Biology* **51**: 49-62.

Lécuyer, Bernard-Pierre. 1987. "Probability in Vital and Social Statistics: Quetelet, Farr, and the Bertillons." In *The Probabilistic Revolution*. Vol. 1, *Ideas in History*, edited by Lorenz Krüger, Lorraine J. Daston, and Michael Heidelberger, 317-335. Cambridge: The MIT Press.

Lerner, I. Michael. [1954] 1970. *Genetic Homeostasis*. Reprint, New York: Dover Publications.

Lewontin, R. C. 1974a. *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.

Lewontin, R. C. 1974b. "The Analysis of Variance and the Analysis of Causes." *American Journal of Human Genetics* **26**: 400-411.

Lewontin, R. C. 1987. "Polymorphism and Heterosis: Old Wine in New Bottles and Vice Versa." *Journal of the History of Biology* **20**: 337-349.

Lewontin, R. C. 1993. *The Doctrine of DNA: Biology as Ideology*. London: Penguin Books.

Lewontin, R. C. and Hubby, J. L. 1966. "A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of Drosophila pseudoobscura." *Genetics* **54**: 595-609.

Limoges, Camille. 1994. "*Errare Humanum Est*: Do Genetic Errors Have a Future?" In *Are Genes Us?: The Social Consequences of the New Genetics*, edited by Carl F. Cranor, 113-124. New Brunswick, N.J.: Rutgers University Press.

Lipmann, Abby. 1991. "Prenatal Genetic Testing and Screening: Constructing Needs and Reinforcing Inequities." *American Journal of Law and Medicine* **42**: 15-50.

Little, Peter. 1990. "Clone maps made simple." *Nature* **346** (16 August): 611-612.

Luria, S. E. 1989. "Human Genome Project," letter to editor. *Science* **246** (17 November): 873.

Macer, Darryl. 1991. "Whose Genome Project?" *Bioethics* **5**: 183-211.

Macilwain, Colin. 1997a. "Human genome centre receives a leg-up." *Nature* **385** (23 January): 283.

Macilwain, Colin. 1997b. "Diversity Project 'Does Not Merit Federal Funding.'" *Nature* **389** (23 October): 774.

Mackie, J. L. 1965. "Causes and Conditions." *American Philosophical Quarterly* **2**: 245-264.

Mackie, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford: Oxford University Press.

Maddox, John. 1991. "The case for the human genome." *Nature* **352** (4 July): 11-14.

Marshall, Eliot. 1997a. "The Battle Over *BRCA1* Goes to Court; *BRCA2* May Be Next." *Science* **278** (12 December): 1874.

Marshall, Eliot. 1997b. "'Playing Chicken' Over Gene Markers." *Science* **278** (19 December): 2046-2048.

Marshall, Eliot, and Elizabeth Pennisi. 1996. "NIH Launches the Final Push to Sequence the Genome." *Science* **272** (12 April): 188-189.

Marx, Jean. 1993. "Genome Project Plans Described." *Science* **260** (9 April): 152-153.

Marzuki, S., et al. 1991. "Normal Variants of Human Mitochondrial DNA and Translation Products: The Building of a Reference Data Base." *Human Genetics* **88**: 139-145.

Marzuki, Sangkot, et al. 1992. "Reply to Howell et al.: The Need for a Joint Effort in the Construction of a Reference Data Base for Normal Sequence Variants of Human mtDNA." *American Journal of Human Genetics* **50**: 1337-1340.

Mauron, Alex, and Jean-Marie Thévoz. 1991. "Germ-line Engineering: A Few European Voices." *Journal of Medicine and Philosophy* **16**: 649-666.

Mayr, Ernst. 1959. "Where Are We?" *Cold Spring Harbor Symposia on Quantitative Biology* **24**: 1-14.

Mayr, Ernst. 1961. "Cause and Effect in Biology." *Science* **134** (10 November): 1501-1506.

Mayr, Ernst. 1963. *Animal Species and Evolution*. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Mayr, Ernst. 1984. "Typological versus Population Thinking." In *Conceptual Issues in Evolutionary Biology: An Anthology*, edited by Elliott Sober, 14-17. Cambridge, Mass.: MIT Press. Originally published in Ernst Mayr, *Evolution and the Diversity of Life* (Cambridge, Mass.: The Belknap Press of Harvard University Press, 1976).

McKusick, Victor A. 1986. "The Gene Map of *Homo sapiens*: Status and Prospectus." *Cold Spring Harbor Symposia on Quantitative Biology* **51**: 15-27.

McKusick, Victor A. 1989. "Mapping and Sequencing the Human Genome." *The New England Journal of Medicine* **320** (6 April): 910-915.

McKusick, Victor A. 1992. *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-Linked Phenotypes*, 2 vols. 10th ed. Baltimore: The Johns Hopkins University Press.

Milkman, Roger. 1967. "Heterosis as a Major Cause of Heterozygosity in Nature." *Genetics* **55**: 493-495.

Milkman, Roger. 1985. "Two Elements of a Unified Theory of Population Genetics and Molecular Evolution." In *Population Genetics and Molecular Evolution: Papers Marking the Sixtieth Birthday of Motoo Kimura*, edited by Tomoko Ohta and Kenichi Aoki, 65-83. Tokyo: Japan Scientific Societies Press.

Millikan, Ruth Garrett. 1989a. "In Defense of Proper Functions." *Philosophy of Science* **56**: 288-302.

Millikan, Ruth Garrett. 1989b. "An Ambiguity in the Notion 'Function.'" *Biology and Philosophy* **4**: 172-176.

Morgan, T. H. 1910. "Sex Limited Inheritance in Drosophila." *Science* **32** (22 July): 120-122.

Muller, H. J. 1922. "Variation Due to Change in the Individual Gene." *The American Naturalist* **56**: 48-49.

Muller, H. J. 1950a. "Evidence of the Precision of Genetic Adaptation." *The Harvey Lectures, Ser. 63, 1947-48*: 165-229.

Muller, H. J. 1950b. "Our Load of Mutations." *The American Journal of Human Genetics* **2**: 111-176.

Muller, H. J. 1973a. "The Darwinian and Modern Conceptions of Natural Selection." In *The Modern Concept of Nature: Essays on Theoretical Biology and Evolution by H.J. Muller*, edited by Elof Axel Carlson, 176-200. Albany: State University of New York Press. First published in *Proceedings of the American Philosophical Society* **93** (1949): 459-470.

Muller, H. J. 1973b. "Evolution by Mutation." In *The Modern Concept of Nature: Essays on Theoretical Biology and Evolution*, edited by Elof Axel Carlson, 201-229. Albany: State University of New York Press, 1973b. First published in *Bulletin of the American Mathematical Society* **64** (1958): 137-160.

Muller, H. J. 1973c. "The Radiation Danger." In *Man's Future Birthright: Essays on Science and Humanity*, edited by Elof Axel Carlson. Albany: State University of New York Press. First published in *The Colorado Quarterly* **6** (Winter 1958): 229-254.

Mulvihill, E. R., J. P. Lepennec and P. Chambon. 1982. "Chicken Oviduct Progesterone Receptor: Location of Specific Regions of High Affinity Binding in Cloned DNA Fragments of Hormone-Responsive Genes." *Cell* **28**: 621-632.

Murphy, Timothy F. 1994. "The Genome Project and the Meaning of Difference." In *Justice and the Human Genome Project*, edited by Timothy F. Murphy and Marc Lappé, 1-13. Berkeley: University of California Press.

Nakagawa, Yusuke, et al. 1995. "A New Mitochondrial DNA Mutation Associated with Non-Insulin-Dependent Diabetes Mellitus." *Biochemical and Biophysical Research Communications* **209**: 664-668.

Navre, M., and H. K. Schachman. 1983. "Synthesis of Aspartate Transcarbamoylase in Escherichia coli: Transcriptional Regulation of the pyrB-pyrI Operon." *Proceedings of the National Academy of Sciences of the United States of America* **80**: 1207-1211.

Neander, Karen. 1991. "Functions as Selected Effects: The Conceptual Analyst's Defense." *Philosophy of Science* **58**: 168-184.

Neilan, B. A., et al. 1994. "Microsatellite Genome Screening: Rapid Non-Denaturing, Non-Isotopic Dinucleotide Repeat Analysis." *Biotechniques* **17** (4): 708, 710, 712.

Nelkin, Dorothy, and Laurence Tancredi. 1989. *Dangerous Diagnostics: The Social Power of Biological Information*. New York: Basic Books.

Nesse, Randolph M., and George C. Williams. 1994. *Why We Get Sick: The New Science of Darwinian Medicine*. New York: Times Books.

Nishino, Ichizo, et al. 1996. "The 3260 Mutation in Mitochondrial DNA Can Cause Mitochondrial Myopathy, Encephalopathy, Lactic Acidosis, and Strokelike Episodes (MELAS)." *Muscle & Nerve* **19**: 1603-1604.

Nordenfelt, Lennart. 1981. *Causation: An Essay*. Stockholm: Forlaget Akademilitteratur.

Nordenfelt, Lennart. 1993. "Concepts of Health and Their Consequences for Health Care." *Theoretical Medicine* **14**: 277-285.

Ohta, Tomoko. 1973. "Slightly deleterious mutant substitutions in evolution." *Nature* **246**: 96-98.

Oxford English Dictionary (shorter version). [1944] last revised 1975. 3rd ed. Oxford: Clarendon Press.

Ozawa, Takayuki. 1995. "Mechanism of Somatic Mitochondrial DNA Mutations Associated with Age and Diseases." *Biochimica et Biophysica Acta* **1271**: 177-189.

Paterson, Andrew H., ed. 1998. *Molecular Dissection of Complex Traits*. Boca Raton: CRC Press.

Paul, Diane B. 1994a. "Eugenic Anxieties, Social Realities, and Political Choices." In *Are Genes Us?: The Social Consequences of the New Genetics*, edited by Carl F. Cranor, 142-154. New Brunswick, N.J.: Rutgers University Press.

Paul, Diane B. 1994b. "Is Human Genetics Disguised Eugenics?" In *Genes and Human Self-Knowledge: Historical and Philosophical Reflections on Modern Genetics*, edited by Robert F. Weir, Susan C. Lawrence, and Evan Fales, 67-83. Iowa City: University of Iowa Press.

Paul, Diane B. 1995. *Controlling Human Heredity: 1865 to the Present*. Atlantic Highlands, New Jersey: Humanities Press.

Plomin, Robert, J. C. DeFries, and G. E. McClearn. 1990. *Behavioral Genetics: A Primer.* 2nd ed. New York: W. H. Freeman and Company.

Plomin, Robert, and Scott Hershberger. 1991. "Genotype-Environment Interaction." In *Conceptualization and Measurement of Organism-Environment Interaction*, edited by Theodore D. Wachs and Robert Plomin, 29-43. Washington, D.C.: American Psychological Association.

Pörn, Ingmar. 1993. "Health and Adaptedness." *Theoretical Medicine* **14**: 295-303.

Porter, Theodore M. 1986. *The Rise of Statistical Thinking, 1820-1900.* Princeton: Princeton University Press.

Porter, Theodore M. 1987. *Lawless Society: Social Science and the Reinterpretation of Statistics in Germany, 1850-1880.* In *The Probabilistic Revolution.* Vol. 1, *Ideas in History*, edited by Lorenz Krüger, Lorraine J. Daston, and Michael Heidelberger, 351-375. Cambridge: The MIT Press.

Provine, William B. 1986. *Sewall Wright and Evolutionary Biology.* Chicago: University of Chicago Press.

Quetelet, Lambert A. J. [1842] 1969. *A Treatise on Man and the Development of his Faculties.* Reprint, with an introduction by Solomon Diamond, Gainsville: Scholars' Facsimiles & Reprints. Originally published in 1835.

Ridley, Mark. 1993. *Evolution.* Boston: Blackwell Scientific Publications.

Rix, Bo Andreassen. 1991. "Should Ethical Concerns Regulate Science?: The European Experience with the Human Genome Project." *Bioethics* **5**: 250-256.

Roberts, Leslie. 1993. "Taking Stock of the Genome Project." *Science* **262** (1 October): 20-22.

Rogers, J. and R. Wall. 1980. *Proceedings of the National Academy of Sciences of the United States of America* **77**: 1877-1879.

Ross, Dennis W. 1992. *Introduction to Molecular Medicine.* New York: Springer-Verlag.

Roush, Wade. 1997. "Whitehead, Three Firms Splice a Deal." *Science* **276** (9 May): 891.

Rowen, Lee, Gregory Mahairas, and Leroy Hood. 1997. "Sequencing the Human Genome." *Science* **278** (24 October): 605-607.

Rusconi, Sandro. 1996. "Transgenic Regulation in Laboratory Animals." In *Molecular and Cellular Genetics*, edited by E. Edward Bittar and Neville Bitter, 377-401. Greenwich, CT: JAI Press.

Ruvolo, Maryellen, et al. 1991. "Resolution of the African Hominid Trichotomy by Use of a Mitochondrial Gene Sequence." *Proceedings of the National Academy of Sciences USA* **88**: 1570-1574.

Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton, N.J.: Princeton University Press.

Sarkar, Sahotra. 1996. "Biological Information: A Skeptical Look at Some Central Dogmas of Molecular Biology." In *The Philosophy and History of Molecular Biology: New Perspectives*, edited by Sahotra Sarkar, 187-231. Dordrecht: Kluwer.

Schmalhausen, I. I. 1949. *Factors of Evolution: The Theory of Stabilizing Selection.* Translated by Isadore Dordick and edited by Theodosius Dobzhansky. Philadelphia: The Blakiston Company.

Schuler, G. D., et al. 1996. "A Gene Map of the Human Genome." *Science* **274** (25 October): 540-546.

Smith, Kelly C. 1992. "The New Problem of Genetics: A Response to Gifford." *Biology and Philosophy* **7**: 331-348.

Sober, Elliott. 1980. "Evolution, Population Thinking, and Essentialism." *Philosophy of Science* **47**: 350-383.

Sober, Elliott. 1984a. *The Nature of Selection.* Cambridge, Mass.: MIT Press.

Sober, Elliott, ed. 1984b. *Conceptual Issues in Evolutionary Biology: An Anthology.* Cambridge, Mass.: MIT Press.

Sterelny, Kim, and Philip Kitcher. 1988. "The Return of the Gene." *The Journal of Philosophy* **85**: 339-361.

Strohman, Richard C. 1993. "Ancient Genomes, Wise Bodies, Unhealthy People: Limits of a Genetic Paradigm in Biology and Medicine." *Perspectives in Biology and Medicine* **37**: 112-145.

Stoneking, Mark, and Terry Melton. 1995. "Establishing the Identity of Anna Anderson Manahan." *Nature Genetics* **9**: 9-10.

Sturtevant, A. H. 1965. *A History of Genetics.* New York: Harper & Row.

Suppes, Patrick. 1970. *A Probabilistic Theory of Causation.* Amsterdam: North Holland.

Sved, J. A., T. E. Reed, and W. F. Bodmer. 1967. "The Number of Balanced Polymorphisms that can be Maintained in a Natural Population." *Genetics* **55**: 469-481.

Sykes, Bryan, et al. 1995. "The Origins of the Polynesians: An Interpretation from Mitochondrial Lineage Analysis." *American Journal of Human Genetics* **57**: 1463-1475.

Tajima, Fumio. 1997. "Estimation of the Amount of DNA Polymorphism and Statistical Tests of the Neutral Mutation Hypothesis Based on DNA Polymorphism." In *Progress in Population Genetics and Human Evolution*, edited by Peter Donnelly and Simon Tavaré, 149-164. New York: Springer.

Takahata, Naoyuki, ed. 1994. *Population Genetics, Molecular Evolution, and the Neutral Theory: Selected Papers of Motoo Kimura*. Chicago: University of Chicago Press.

Thompson, Larry. 1993. "Healy and Collins Strike a Deal." *Science* **259** (1 January): 22-23.

Torroni, Antonio, and Douglas C. Wallace. 1994. "Mitochondrial DNA Variation in Human Populations and Implications for Detection of Mitochondrial DNA Mutations of Pathological Significance." *Journal of Bioenergetics and Biomembranes* **26**: 261-271.

van Fraassen, Bas. 1980. *The Scientific Image*. Oxford: Clarendon Press.

Wachs, Theodore D. 1991. "Synthesis: Promising Research Designs, Measures, and Strategies." In *Conceptualization and Measurement of Organism-Environment Interaction*, edited by Theodore D. Wachs and Robert Plomin, 162-182. Washington, D.C.: American Psychological Association, Washington, D.C.

Waddington, C. H. 1957. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. London: Allen R. Unwin.

Wadman, Meredith. 1998. "Sequencing funds 'must remain focused.'" *Nature* **391** (8 January): 114-115.

Wainscoat, Jim. 1987. "Out of the garden of Eden." *Nature* **325**: 13.

Wallace, Bruce, and Theodosius Dobzhansky. 1959. *Radiation, Genes, and Man*. New York: Holt, Rinehart, and Winston.

Wallace, Bruce. 1991. *Fifty Years of Genetic Load: An Odyssey*. Ithaca: Cornell University Press.

Ward, R.H., et al. 1991. "Extensive Mitochondrial Diversity within a Single Amerindian Tribe." *Proceedings of the National Academy of Sciences USA* **88**: 8720-8724.

Watson, James D. 1990. "The Human Genome Project: Past, Present, and Future." *Science* **248** (6 April): 44-49.

Watson, James D. 1992. "A Personal View of the Project." In *The Code of Codes: Scientific and Social Issues in the Human Genome Project*, edited by Daniel J. Kevles and Leroy Hood, 164-173. Cambridge, Mass.: Harvard University Press.

Watson, James D. 1993. "Looking Forward." *Gene* **135**: 309-315.

Watson, James D., and Robert Mullan Cook-Deegan. 1990. "The Human Genome Project and International Health." *JAMA* **263** (27 June): 3322-3324.

Weatherall, David. 1994. "The Relative Roles of Nature and Nurture in Common Disease." In *Health and the Environment: The Linacre Lectures, 1992-3*, edited by Bryan Cartledge, 172-198. Oxford: Oxford University Press.

Weaver, C.A., D.F. Gordon, and B. Kemper. 1982. "Nucleotide Sequence of Bovine Parathyroid Hormone Messenger RNA." *Molecular and Cellular Endocrinology* **28**: 411-424.

Weiss, Kenneth M. 1996a. "Introduction." In *Variation in the Human Genome*, edited by Derek Chadwick and Gail Cardew, 1-5. West Sussex: John Wiley & Sons.

Weiss, Kenneth M. 1996b. "Biological Anthropology." *Nature* **379** (25 January): 293.

Wexler, I. D., et al. 1988. "Heterogeneous Expression of Protein and mRNA in Pyruvate Dehydrogenase Deficiency." *Proceedings of the National Academy of Sciences USA* **85**: 7336-7340.

Whitbeck, Caroline. 1981. "A Theory of Health." In *Concepts of Health and Disease: Interdisciplinary Perspectives*, edited by Arthur L. Caplan, H. Tristram Engelhardt, Jr., and James J. McCartney. London: Addison-Wesley Publishing Co.

White, R., et al. 1986. "Construction of Human Genetic Linkage Maps: I. Progress and Perspectives." *Cold Spring Harbor Symposia on Quantitative Biology* **51**: 29-38.

Wilkie, Tom. 1994. *Perilous Knowledge: The Human Genome and Its Implications*. London: Faber and Faber.

World Health Organization. 1981. "Constitution of the World Health Organization." In *Concepts of Health and Disease: Interdisciplinary Perspectives*, edited by Arthur L. Caplan, H. Tristram Engelhardt, Jr., and James J. McCartney, 83-84. London: Addison-Wesley Publishing Co. Constitution adopted in 1946 and first published in *The First Ten Years of the World Health Organization* (Geneva: WHO, 1958).

Wright, Larry. 1984. "Functions." In *Conceptual Issues in Evolutionary Biology: An Anthology*, edited by Elliot Sober, 347-368. Cambridge, Mass.: The MIT Press. First published in *Philosophical Review* **82** (1973): 139-168.

Wulff, Henrik R. 1984. "Comments on Hesslow's 'What is a Genetic Disease?'" In *Health, Disease, and Causal Explanations in Medicine*, edited by Lennart Nordenfelt and B. Ingemar B. Lindahl, 195-197. Dordrecht: D. Reidel.

Wyndham, John. 1958. *The Chrysalids*. Harmondsmith: Penguin Books.

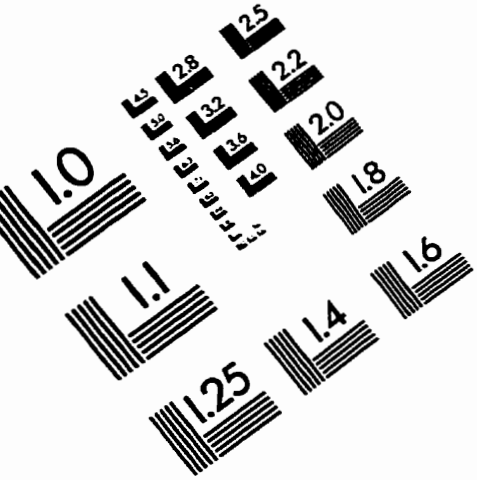
Xu, Xiufeng, and Ulfur Arnason. 1996. "A Complete Sequence of the Mitochondrial Genome of the Western Lowland Gorilla." *Molecular Biology and Evolution* **13**: 691-698.

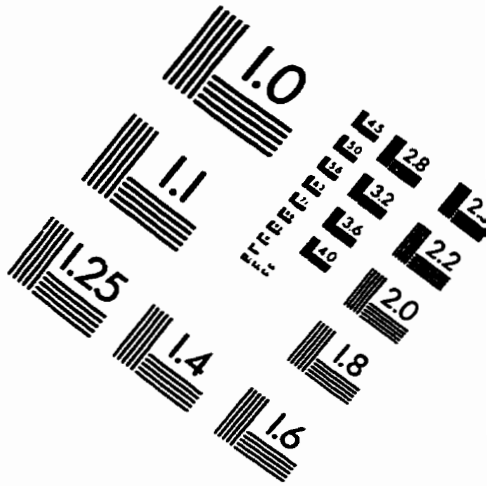
Yoxen, E.J. 1984. "Constructing Genetic Diseases." In *Cultural Perspectives on Biological Knowledge*, edited by Troy Duster and Karen Garett, 41-62. Norwood, N.J.: Ablex.

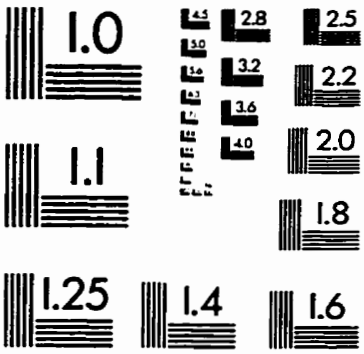
Zimmerman, Burke K. 1991. "Human Germ-line Therapy: The Case for its Development and Use." *Journal of Medicine and Philosophy* **16**: 593-612.

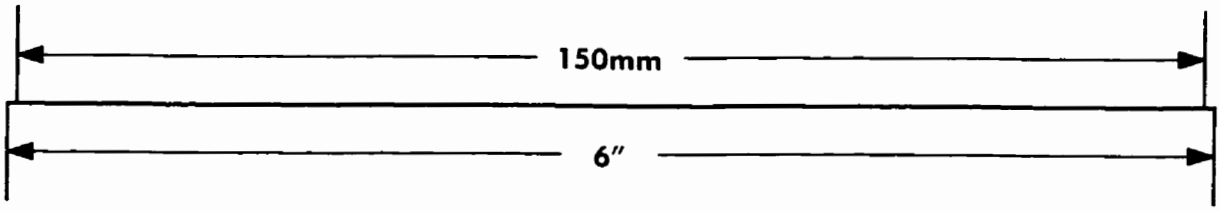
Zyskind, J.W., et al. 1983. "Chromosomal Replication Origin from the Marine Bacterium Vibrio Harveyi Functions in Escherichia coli: oriC consensus sequence." *Proceedings of the National Academy of Sciences USA* **80**: 1164-1168.

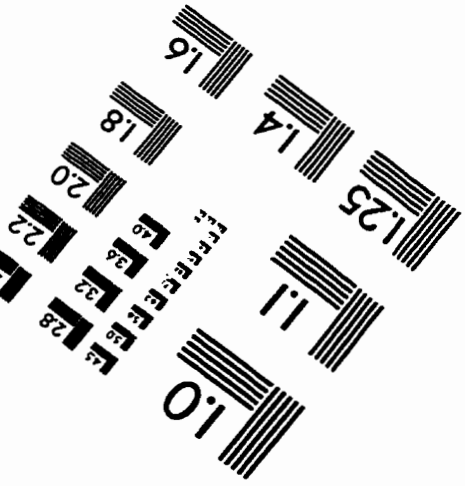
IMAGE EVALUATION TEST TARGET (QA-3)

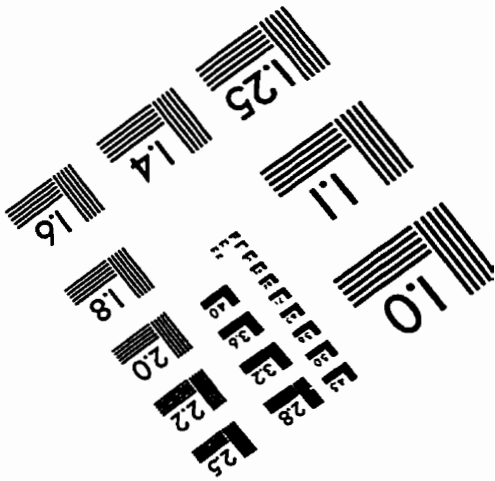












APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989