Prolegomena: Genes, Science, and Science Fiction

If one looks at mass media headlines, one will find several accounts of how genes determine various aspects of our lives. Many of these claim to take into account conclusions from recent research in genetics. The general impression is that there exist "genes for" characters,[1] i.e. that single genes cause even complex characters. This view seems to be quite prevalent e.g. it is common to find teachers teaching that genes determine characters, media reports presenting studies that found associations between particular genes and particular diseases, and personal observations of the development of characters that do not seem to be affected by the environment (Moore, 2008). A quick search on the World Wide Web reveals several examples. For instance, a 2014 article in the *Guardian* was titled "Happy gene' may increase chances of romantic relationships."[2] The title of a 2015 article in the *New York Times* suggested that "Infidelity lurks in your genes."[3] A 2014 article in *Time* magazine was titled: "The genes responsible for deadly prostate cancer discovered."[4] And there are more. Several authors have argued that messages like these impose genetic determinist views on the public (e.g. Hubbard & Wald, 1997; Nelkin & Lindee, 2004). This certainly seems plausible, particularly as many people might just read the headlines such as those mentioned previously, without ever reading the full article that might suggest otherwise. Therefore, they might conclude that genes determine who we are.

The problem of making sense of genes, i.e. understanding what genes are and what they do, has concerned me a lot and for a long time. However,

[1] To avoid inconsistencies while referring to features, traits, characteristics, and so on interchangeably, I am using the term "character" throughout this book, which can be defined as any recognizable feature of an organism that can exist in a variety of character states, and at several levels from the molecular to the organismal (based on Arthur, 2004, p. 212). Disease conditions will be considered as character states that deviate from what we tend to consider as "normal."

[2] www.theguardian.com/science/2014/nov/20/happy-gene-romantic-relationship-serotonin-romance

[3] www.nytimes.com/2015/05/24/opinion/sunday/infidelity-lurks-in-your-genes.html?partner=rss&emc=rss

[4] http://time.com/96247/scientists-have-discovered-the-two-genes-responsible-for-aggressive-prostate-cancer/

in my previous book, *Understanding Evolution* (Kampourakis, 2014), I refrained from using the term gene at all. Instead, I referred to genetic material and DNA sequences that are implicated in biological phenomena. Eventually, it was possible to write a whole book without any reference to genes. Yet, ignoring the problem does not contribute anything to its solution, and so I decided to devote my second book to the gene concept that was put aside in my first one. There are two reasons for this. On the one hand, the term exists in the public discourse and so it is better to try to clarify it rather than just ignore it. On the other hand, scientists use the term in their work and in its public presentation. Therefore, I thought that I could make a minor contribution to countering the public distortions of the gene concept and help students in the life sciences, biologists, biology teachers, health professionals, and anyone else interested in acquiring a better understanding of it, as well as provide them with conceptual tools to explain genes to nonexperts.

Generally speaking, our knowledge takes the form of concepts that are mental representations of the world. Concepts should be distinguished from conceptions, the latter being the different meanings of, or meanings associated with, particular concepts. This means that whereas we may generally agree on a general definition of a certain concept, e.g. "dog," people all over the world may hold different conceptions of what a dog is or looks like. In other words, even if a concept is well defined and even if it is clear to people to what this concept refers, individual conceptions may vary a lot if one takes the time to consider them. This is also the case for scientific concepts, such as the gene. Scientific concepts are systematic mental representations of the world through which explanations of and predictions about phenomena are possible (Nersessian, 2008, p. 186). In this case, the difference between concepts and conceptions becomes more striking; whereas scientists may agree on the definition of a certain concept, nonexperts may hold very different conceptions of it for various reasons. Such reasons may include the public distortions of the concept under discussion, or that people simply failed to understand it because of their own preconceptions. In the present book I focus on the gene concept that most people have heard of, but many fail to understand. My aim is to explain this concept and address certain prevalent but inaccurate conceptions. At the end of this book, the reader should have acquired a better understanding of what a gene is and is not, as well as what a gene can and cannot do.

You have probably had some genetics classes during your secondary school years. Even if you have forgotten most of what you learned at that time, you probably remember Gregor Mendel (1822–1884) and his experiments with peas. Because of these experiments Mendel is considered as a pioneer of genetics and as the person who discovered the laws of heredity. You may also remember that, according to your secondary school genetics, things were rather simple and straightforward with inherited characters. In the case of disease, for instance, your teacher used to explain to you that most "normal"[5] alleles (the different versions of the same gene) were dominant, i.e. imposed their effects on the pathogenic ones that were therefore recessive. Thus, when an individual had one normal, dominant allele and one pathogenic, recessive allele, there was no problem. However, in some cases it was possible for such individuals to have offspring with the disease, because two recessive pathogenic alleles had come together in the same individual. Why was that? According to your teacher, the normal allele somehow determined a normal character, whereas the pathogenic allele determined a pathogenic version of the same character that occurred because e.g. some important factor was missing. Thus, a person with two pathogenic alleles totally lacked that factor and so had the disease. Genes could thus determine characters and diseases.

So, what was a gene according to your high school genetics? It was usually defined as a segment of DNA that contained the information for the production (or not) of some protein[6] that in turn somehow determined a character. A definition could not be clearer, could it? In definitions like these, the take-home message is usually that genes work in a deterministic if-you-have-the-gene-you-will-also-have-the-character kind of way (Moore, 2013a).[7] However, if one looks more closely at such definitions, one will realize that genes are conceived as simultaneously operating at two levels: the molecular (production of a protein) and the organismal (determination of a character). What is implied is that the molecular level (DNA/gene) somehow determines the organismal. Even

---

[5] Defining what is normal and what is not normal is quite difficult and subjective sometimes. In this book, I will use the term in a rather vague sense to refer to whatever state can be considered as natural and unproblematic.

[6] DNA, or deoxyribonucleic acid, is a long molecule that consists of consecutive nucleotides. Proteins, or polypeptides, are long molecules that consist of consecutive amino acids.

[7] This is the topic of Chapters 6 and 7.

when gene definitions do not simultaneously refer to both levels, one may find distinct definitions referring to the molecular level and the organismal level to co-exist.

Consider, for example, the definitions of gene in two very good textbooks that I used when I was still teaching biology at school. The first of these (Walpole et al., 2011) contained the following definitions in the main text: "A gene is a particular section of a DNA strand that, when transcribed and translated, forms a specific polypeptide" (p. 67), and a "Gene [is] a heritable factor that controls a specific characteristic, or a section of DNA that codes for the formation of a polypeptide" (p. 68). In both cases, the gene is described as a section of DNA, but in the second definition it is also described as a factor that controls a character. However, if one looks at the glossary of the same book, no reference to DNA is made. A gene is defined there as "a heritable factor that controls a specific characteristic" (p. 586). The case is similar in the other textbook (Sadava et al., 2011). In the main text, genes are defined as segments or sequences of DNA that encode proteins: "Genes are specific segments of DNA encoding the information the cell uses to make proteins" (p. 6); "The sequences of DNA that encode specific proteins are transcribed into RNA and are called genes" (p. 64); and "a gene is a sequence of DNA that resides at a particular site on a chromosome, called a locus (plural loci). Genes are expressed in the phenotype mostly as proteins with particular functions, such as enzymes" (p. 242). However, the definition in the glossary of the same book pays more attention to function, ignoring structure. The gene is defined there as: "A unit of heredity. Used here as the unit of genetic function which carries the information for a single polypeptide or RNA" (p. G-12).

Is there a consistency problem here? Why are the definitions in the main text and the glossary of the same textbook different? The definitions of the term "gene" in the glossaries of both textbooks refer to a hereditary factor without specifying what exactly this is made of. In contrast, the definitions in the main text of both textbooks are specific about what is referred to by the term "gene": a section, segment, or sequence of DNA. Are these textbooks referring to the same concept at different levels of organization? In the present book, I show that the definitions in the main texts and in the glossaries of these textbooks are in fact very different. I also explain that the reasons why such different definitions co-exist in the same textbook are not pedagogical

but historical. By presenting how the gene concept was coined and has evolved over the past 100 years or so, during which time research on heredity has been conducted, I show that different gene concepts have dominated discourse on heredity over different periods and that, recently, more than one have co-existed.

The next question that arises is this: What is it that genes do? If you open a newspaper or a popular magazine it is very likely that you will read a report about a recent discovery of a "gene for" something. Genes have been reported to determine characters of all kinds, such as eye color and height. They have also been reported to determine well-studied diseases, such as thalassemia and phenylketonuria, but also more complex and less-well understood ones such as coronary heart disease and cancer. Most interestingly, genes are often reported in the popular press to determine all kinds of behaviors and psychological states. Thus, "genes for" depression, schizophrenia, intelligence, alcoholism, criminality, promiscuity, homosexuality, and more have been reported to exist. As a result, genes are perceived as determining everything. This is particularly evident in characters that run in families, which are, often without a second thought, attributed to genes inherited from parents to offspring, and not to other possible factors such as their shared environment.

I speculate that if there was a report that George H. W. Bush (1924–) and his son George W. Bush (1946–) were both elected presidents of the United States because of a particular gene they both had, perhaps a "gene for" US presidency, many people would not question such a conclusion. Similarly, many people might find reasonable that there exists a "gene for" becoming a Hollywood star in the case of Kirk Douglas (1916–) and his son Michael (1944–), or in the case of Judy Garland (1922–1969) and her daughter Liza Minelli (1946–). These same people might attribute to a "gene for" the Nobel Prize the fact that both Arthur Kornberg (1918–2007) and his son Roger (1947–) were awarded a Nobel Prize – but perhaps different versions of that gene could account for the fact that Arthur's prize was in physiology and medicine, whereas Roger's was in chemistry. These examples might sound exaggerated, but as I show later in this book, claims like these are quite common in the public sphere. For many people, the interesting question is not whether genes determine characters and behaviors; the common assumption is that they do. The interesting question is how they do it.

The metaphors currently used about genes present them as autonomous entities, which both contain all the necessary information to determine characters and are capable of making use of it. Therefore, both in research and in popular parlance, genes have been described as the "essences" of life, as the absolute "determinants" of characters and disease and therefore as providing the ultimate explanations for all biological phenomena because the latter can be "reduced" to the gene level and thus be explained. These views have been described as *genetic essentialism*, *genetic determinism*, and *genetic reductionism*, respectively. They are all related to one another, and they may even seem to overlap. However, they are distinct and should not be confused. In order to avoid confusion and overlaps in definitions, in this book I use the following definitions (based on Beckwith, 2002; Kitcher, 2003; Wilkins, 2013):

- *Genetic essentialism:* genes are fixed entities, which are transferred unchanged across generations and which are the essence of what we are by specifying characters from which their existence can be inferred.
- *Genetic determinism:* genes invariably determine characters, so that the outcomes are just a little, or not at all, affected by changes in the environment or by the different environments in which individuals live.
- *Genetic reductionism:* genes provide the ultimate explanation for characters, and so the best approach to explain these is by studying phenomena at the level of genes.

Most importantly, these are the onerous conceptions that the present book aims at addressing.

Whether or not these conceptions are distinct apparently depends on how one defines them. I use these definitions in order to distinguish between three important properties usually attributed to genes: (1) that they are fixed essences; (2) that they alone determine characters notwithstanding the environment; and (3) that they best explain the presence of characters. The power attributed to genes has often gone beyond the realm of science to reach that of science fiction. Genes have been described as autonomous, self-replicating entities capable of doing everything and of determining everything. There are "fat" genes, "smart" genes, "cancer" genes, "infidelity" genes, "aggression" genes, "happiness" genes, "God" genes, and more (a World Wide Web search of these terms is illuminating;

in some cases, even books with titles like these exist). The underlying assumption in most cases is that much of what we are or do is driven (if not dictated) by our genes. Perhaps we find attributing whatever happens to one's genes very intuitive, because it makes sense immediately? It is the supernatural powers attributed to genes that this book aims at addressing. Of course, I am not going to argue that genes are not important – they are! But it is one thing to say that genes are important for what we are or do, and another that they are the ultimate determinants of these. I hope that, at the end this book, I will have succeeded at clarifying what genes are and are not, as well as what they can and cannot do.

Chapters 1–4 provide a brief account of how the initially "empty," or, to be more precise, referentially indefinite (i.e. that did not refer to a particular entity), gene concept came to have two distinct meanings during the twentieth century: that of a hypothetical inherited factor, the changes in which were somehow related to changes in characters, and that of a DNA sequence that encoded the information for a protein. Whereas it may have initially seemed self-evident that these two gene concepts might overlap and that they would converge to the same segments of DNA, by the 1970s it became quite evident that this is not the case. More recent research has shown that it is impossible to structurally individuate genes, and that the best we can do is to identify them on the basis of their functional products. I must note that in these chapters I do not intend to provide a detailed and complete history of the "gene" concept (for such histories see Beurton et al., 2000; Falk 2009; Rheinberger et al., 2015; Rheinberger & Müller-Wille, in press). Rather, these chapters aim at providing an idea of the complexities of precisely defining what a gene is.

Then, in Chapters 5–8, I describe the presentations of genes in the media and on the websites of companies selling genetic tests. I show that the underlying message in many cases is that there are genes that determine characters and disease. I also present research on the conceptions that students and the public hold about genes and the difficulties they face in understanding what genes are and do. Then, I show that simple, causal connections between genes and characters or genes and disease are not adequate to accurately represent the actual phenomena. Research in genetics shows that these are actually very complicated. In many cases, single genes cannot explain the variation observed not only for complex characters and disease but also for simple monogenic ones. On the basis of these, I conclude by explaining that genes do not actually

do anything on their own. I also explain why the notion of "genes for," in the vernacular sense, is not only misleading but also entirely inaccurate and scientifically illegitimate.

Finally, in Chapters 9–12, I come to some major conclusions from the research presented in the previous chapters. First, I show that genes "operate" in the context of developmental processes only. This means that genes are implicated in the development of characters but do not determine them. Second, I explain why single genes do not alone produce characters or disease but contribute to their variation. This means that genes can account for variation in characters but cannot alone explain their origin. Third, I show that genes are not the masters of the game but are subject to complex regulatory processes. There seem to exist many regulatory sequences in what until recently has been called "junk" DNA. As a result, the genome of an organism is more than the sum of its genes. Finally, I discuss in some detail the limitations of genetic testing that are not often taken into account in public discourse, in order to show what is and what is not currently possible to achieve from DNA analyses, and to debunk the myth of their infallibility. I also show how misleading information about genes can be when it comes to probabilistic thinking.

The chapters of this book could be read independently from one another; however, in many cases individual chapters build on knowledge and understanding of concepts that have been presented in previous ones. Therefore, I recommend that you read this book from beginning to end, without skipping any chapters – unless you are very well familiar with the respective topics. However, for those readers who decide not to do so, the book includes a glossary with the definitions of the most important concepts. Next to that, there is also a guide to further reading that includes relevant books that treat in more detail many of the topics presented in this book. In most cases, I have read and cited the original research articles. However, in several cases I found the accounts given in certain books useful or the ideas illuminating, and so I am citing these. Many of the topics I present are discussed in several books, but I am only citing them wherever it is really useful. The *Further Reading* section provides information about the books one should read after reading the present one.

A central feature of the present book is that it is mostly about human characters and disease. When this is not the case, it is usually about phenomena of relevance to human life. I must note that this is not due to any anthropocentricism on my part. Quite the contrary, I believe that we are not anything special in this world, or at least that we are not any

more special than any other organism that lives in it. Nevertheless, I thought that the book would be more interesting to readers if I discussed phenomena about, or relevant to, human life. I made the decision to focus on human genes because my experience as a teacher and an educator was that students' interest was aroused whenever a topic about human life or health came up. Pragmatically thinking, making sense of genes also has an important medical interest; therefore, I wanted this book to be useful not only for biologists but also for physicians and other health professionals. This approach is biased, of course, because it overlooks important aspects of life on Earth. But it is also more interesting for humans. I do hope that readers will appreciate both this decision and the outcome. I hope that they will find this biased-toward-humans book interesting and didactic. But they should also keep its bias in mind and avoid unwarranted generalizations from the mostly medical-centered and human-focused research presented in this book.

I must also note that the term "genetics" is used throughout the book in a very broad sense to refer to any research about genes. Therefore, the term "genetics" encompasses research in classical genetics of the first half of the twentieth century, molecular biology and genetics of the latter half of the twentieth century, and genomics of the past twenty-five years or so, despite the important differences among these research approaches. Conceptually, "genetics" could be perceived to refer to genes only, whereas "genomics" could be perceived to refer to the genome as a whole, including genes and everything else in DNA. Therefore, "genomics" could be considered as a broader term than "genetics," as the genome is a broader concept than the gene (see Annas & Elias, 2015, p. 3). Nevertheless, as genes have been the main focus of research so far, it is conceptually sound and certainly simple for the purpose of the present book to use the term "genetics" to refer to all research about, or relevant to, genes – no matter how these are defined – also encompassing genomics research.

The present book is intended primarily for non-experts, i.e. people not working on genetics, who want an introduction to genes. The intended audience includes undergraduate students in biology, medicine, and pharmacy, as well as biology teachers and educators. The book provides an overview of the core concepts and issues in genetics, and it can also serve as an introduction to more detailed and advanced forays in the literature. Physicians and other healthcare professionals who are interested in getting a concise overview of contemporary genetics research and

concepts would also find this book useful, while it could also be useful to researchers in biology who are interested in the relationship between biological research and society at large. Finally, the book is appropriate for any lay reader who wants an accessible but rigorous introduction to genes.

I should note that the present book focuses on a research area that advances at an extremely fast pace. I found myself revising and updating the text several times during the one year or so that it took me to write this book. Therefore, I am certain that as soon as the book will be published there will be new research articles that I might have considered, should they have been published before that. However, I think that the main points of the present book and its conceptual foundations will remain unchanged for several years to come, even if we get to know more and understand the respective phenomena in more detail.

A note of caution. Throughout the book, I have used expressions such as "research has shown" or "evidence suggests," etc. Research produces data that becomes evidence within a certain theoretical framework. Neither research, nor evidence "show" or "suggest" anything on their own. Everything in science is a matter of interpretation. Nevertheless, because it would be strange for me to write each time that "research findings have formed the basis for the conclusion that…" or that "evidence has been interpreted as showing that…", I have used expressions like "research has shown" or "evidence suggests" as a kind of shorthand. Nevertheless, you should be aware that these statements have problems and which is the actual meaning behind them (I discuss this topic in more detail in Chapter 12).

Before we proceed, some nomenclature is necessary. Most genes mentioned in this book are human ones. Both gene and protein symbols in humans are written with uppercase letters. However, gene symbols are written with inclined characters, whereas protein symbols are written with regular characters. For example, the symbol for the protein called "monoamine oxidase A" is MAOA, whereas the symbol for the respective gene would be *MAOA*. For the purpose of consistency, all gene symbols and gene names in this book are derived from the Human Genome Organization Gene Nomenclature Committee website: www.genenames.org. For the few other species to the genes of which I refer in this book, I simply use the nomenclature used in the respective articles.

Let us now start trying to make sense of genes.