INFERRING TO THE BEST EXPLANATION:

A DECISION THEORETIC APPROACH

ELEONORA CRESTO

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2006

ABSTRACT

INFERRING TO THE BEST EXPLANATION:

A DECISION THEORETIC APPROACH

Eleonora Cresto

In this dissertation I offer an elucidation of the concept of Inference to the Best Explanation (IBE) with the aid of a decision theoretic apparatus. I conceive of IBE as a complex process that consists of two main moments: an abductive stage (named after Charles Peirce's rich notion of abduction), in which the agent proposes explanatory hypotheses for further consideration, and a selection stage, in which the agent picks the best element, or a disjunction of best elements, out of the antecedent batch.

In order to provide a detailed analysis as to how the IBE process goes, in part A I develop a formal epistemic model. The epistemic state of a given agent at a given time is represented by a tuple consisting of a language, a set of "meaningful" sentences, a set of accepted sentences (or beliefs), a set of personal probability functions over the set of meaningful sentences, a set of explanatory arguments, and a set of argumentative patterns. I contend that it is rational to expect shifts in such tuples throughout the course of an IBE process, and not just as a result of its last step. In particular, I submit that the abductive stage may give rise to "structural" changes, which involve shifts in the set of the agent's meaningful statements.

My description of the selection stage, on the other hand, relies on Isaac Levi's brand of cognitive decision theory. Within Levi's general setting, I argue that a

suitable notion of epistemic utility for hypotheses needs to take *epistemic virtues* into account (features such as unification power, or fertility). In addition, I embrace the idea that the overall explanatory force of a given hypothesis should be identified with its epistemic expected utility. Under certain circumstances, the theory recommends that the agent's set of acceptances be enlarged with the most explanatory hypothesis of the lot.

In part B I apply the model to two turning points in the history of genetics: the rise of Mendelism at the beginning of the twentieth century, and Theodore Avery's research during the forties about bacterial transformation and DNA. Compared to other epistemic models, the present account is found to be better suited to deal with complex historical cases, and more sensitive to the nuances of real-life science.

TABLE OF CONTENTS

PART B: TESTING THE MODEL

ACKNOWLEDGMENTS

Many professors, colleagues and friends helped me go through the process of writing this dissertation.

I owe an enormous debt of gratitude to Isaac Levi for his patience, encouragement, and generosity. I hope that the present work can, at least in part, reflect what I have learnt from him.

I also owe a very special debt to Achille Varzi for his support and for numerous discussions and suggestions. I would also like to thank Philip Kitcher, John Collins, Akeel Bilgrami, Horacio Arló Costa, and Eduardo Fermé, who carefully read excerpts of previous versions of this dissertation, and prevented me from making many mistakes.

I am very grateful to the many fellow graduate students of the Philosophy Department at Columbia University who, in addition to stimulating conversations, offered me their friendship and support. At the risk of omitting some names, I would like to mention Ernesto García, Massimo Grassia, Gustavo Ortiz Millan, Gurpreet Rattan, and Dana Tulodziecki. Special thanks are due to John Brunero for proof-reading the final version of my work, and to Romina Padró for hosting me during one of my visits to New York. I would also like to thank my friends and colleagues Eduardo Barrio, Eleonora Orlando, Federico Penelas, and Verónica Tozzi, from the University of Buenos Aires, as well as all the people at the *GAF* (*Grupo de Acción Filosófica*), for providing an excellent working environment since my move to Argentina in 2001. I would like to acknowledge the support of a fellowship from *Fundación Antorchas*, Argentina, to complete my Ph.D.

On a more personal note, I would like to express my gratitude to my friends Ileana Afeltra, Mara Barreto, Stella Rodiño, Bárbara Pra, and Karina Pedace, for their love and patience. Last, but not least, I would like to thank my parents, and my husband Alejandro Cassini for providing endless encouragement and help during very busy years.

To my family

PREFACE

1. Two Cases:

(a) Between 1856 and 1864 Gregor Mendel performed thousands of breeding experiments with plants of the *Pisum* genus. He concluded, among other things, that all constant forms that resulted from the combination of characters put together at the moment of fecundation were represented in equal number in the sexual cells in the hybrids.[1] This was presumably a direct antecedent, albeit in a less clear form, of the thesis that each heritable trait of an organism is represented by two elements, one derived from the female and the other from the male progenitor.

Forty years later, William Bateson would seek to convert the scientific community to Mendelism. Presumably, he acted in this way because he was convinced that Mendel's laws amounted to *the best explanation* for Mendel's experimental results, as well as for many other known observational facts.

In the long run, Bateson was indeed successful – but back in 1902, he faced the strong opposition of Francis Weldon, who discredited Mendel's observations and contended that the evidence available was *best explained* by means of Galton's ancestral law of heredity.

(b) In 1944, Oswald Avery, Colin MacLeod and Maclyn McCarty published a fundamental paper whose subtitle was "Induction of Transformation by a

[1] Mendel (1866).

Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III."[2] A few years earlier, it had been reported that live, non-virulent type II pneumonia germs could be "transformed" into a virulent, type III strand, by the simple procedure of growing a culture of the live, non-virulent type II germ in the presence of dead Pneumococci of the virulent type III form. This led to the important question of what substance could possibly be responsible for the dramatic change. The 1944 paper described a fascinating attempt at finding such a substance.

Through careful and ingenious procedures, Avery and his team obtained a "fibrous material" out of the virulent type III Pneumococci, which was by itself able to produce the transformation. Then they performed many tests to try to find out what the material was, and what it was not. All tests came back negative for protein and RNA, and they came back positive for DNA. Hence, there seemed to be strong reasons to suspect that the transforming principle was DNA. It was still possible, however, that the transformation might have occurred due to "minute amounts of some other substance adsorbed to [the nucleic acid] or so intimately associated with it as to escape detection."[3]

In the years that followed Avery's experiments, some of his colleagues fully accepted the idea that DNA was the transforming principle; they believed that it was *the best available explanation* for the

[2] Avery *et al.* (1944).

[3] Avery *et al.* (1944), p. 190.

observed results. Others rejected it. A different, but extremely important question was whether the transforming principle (be it DNA or something else) had to be conceptualized as a *gene*. In 1944 Avery refused to address this question, but others were not so cautious.

2. In Search of a Theoretic Apparatus

Although the two cases differ in a number of crucial aspects, I submit that they both exemplify variants of a process that may be called, quite generally, "inference to the best explanation." My goal in this dissertation is to offer a theoretical apparatus that may help to explain what was going on in both Mendel's and Avery's case. The theory must be able to give a rationale for the different possible outcomes mentioned in the previous paragraphs. More generally, in this dissertation I want to offer a rigorous, normative model for the task of inferring to the best explanation. The label "inference to the best explanation" is extensively used in the current philosophical literature, but still lacks a homogeneous interpretation. An ideal elucidation should help us understand past cases of such inferences, and should also have prescriptive power for future cases.

In particular, I shall pay special attention to the way an inference to the best explanation is able to generate epistemic changes, and I shall try to be sensitive to the *types* of change that it may generate. Thus, for example, I will argue in favor of making a basic distinction between radical, "structural" changes, and more standard ones. I shall also seek to explain why it might be rational for different agents that share the same information to arrive at different conclusions concerning the legitimacy of accepting, rejecting, or suspending judgment on a given hypothesis.

I shall divide the dissertation in two parts. Part A comprises chapters 1 to 3: in chapter 1 I discuss possible historical roots of the concept of inference to the best explanation, and review the current literature on the topic; chapters 2 and 3 deal with the construction of the model. In part B I seek to test the model of part A through concrete applications. Chapter 4 presents a brief comparison between Mendel's and Avery´s case; chapter 5 deals with the rise of Mendelism at the beginning of the twentieth century, with an emphasis on the Bateson-Weldon debate, whereas chapter 6 addresses the debate between Avery's team and other biologists of the time on the nature of DNA.