

ELIMINATIVE ABDUCTION EXAMPLES FROM MEDICINE

Forthcoming in *Studies in History and Philosophy of Science*

Alexander Bird

Abstract

Peter Lipton argues that inference to the best explanation (IBE) involves the selection of a hypothesis on the basis of its loveliness. I argue that in optimal cases of IBE we may be able to eliminate all but one of the hypotheses. In such cases we have a form of eliminative induction takes place, which I call 'Holmesian inference'. I argue that Lipton's example in which Ignaz Semmelweis identified a cause of puerperal fever better illustrates Holmesian inference than Liptonian IBE. I consider in detail the conditions under which Holmesian inference is possible and conclude by considering the epistemological relations between Holmesian inference and Liptonian IBE.

keywords inference to the best explanation, Peter Lipton, abduction; Holmesian inference; eliminative induction.

1 Introduction

Many, probably most, scientific realists believe that inference to the best explanation (IBE), broadly construed, is at the heart of science. There is a plethora of techniques, methods, rules of thumb, heuristics and so forth that are used to generate scientific knowledge and which do not fit the IBE mould. Nonetheless, many of our most interesting theoretical discoveries have been made with the application of IBE, including our discoveries concerning unobservable entities and processes.

In the light of this, it is an extraordinary achievement that Peter Lipton has given us *the* authoritative account of what IBE is and how it contributes to theory choice and confirmation. Given the capacity of philosophers for disagreement and for generating theories, one might have thought that Lipton might have had a number of rivals concerning this absolutely crucial topic. But that just does not seem to be the case. Lipton's *Inference to the Best Explanation* seems truly to be a Kuhnian paradigm in the philosophy of science. For those working on inference to the best explanation, it is the text that sets our agenda, that lays out the problems we must contend with, and which, in many dimensions, is an exemplar of how we should carry out our work. In this paper I pursue some normal (philosophy of) science in the Liptonian tradition. While not seeking any revolutionary change to that paradigm, I do want to suggest that there is an important respect in which Lipton's picture of IBE needs supplementing.

I will start by outlining Lipton's conception of IBE. I'll then mention an anomaly that arises in his discussion of his central illustrative case, that of Ignaz Semmelweis and puerperal fever. Outlining the relevant features of that case, and of another case, the discovery of the cause of AIDS, will give an indication of why that anomaly arises, and what supplement to Liptonian IBE is thereby required. That supplement states that in some cases of IBE our evidence permits us to select just one potential explanation as *the* explanation, because it is the *only* potential explanation consistent with the evidence. This I call Holmesian Inference.

2 Inference to the best explanation

Inference to the best explanation is about choosing among explanations. It is a matter of choosing among potential explanations of some phenomenon the one that is the best by certain criteria. If there is a suitable best explanation, IBE says that we may infer that it is the *actual* explanation. If some hypothesis provides the actual explanation of a phenomenon, then that hypothesis is true.

How do we choose among potential explanations? According to Lipton, IBE is a two-stage process, where both stages are filters of potential explanations (?: 56–64):

Stage 1: The first stage filters out the implausible explanations. The imaginative capacity of scientists generates all the plausible potential explanations and just leaves the remainder unconsidered.

Stage 2: At the second stage, scientists investigate the live potential explanations that have passed through the first filter, and ultimately rank them according to their explanatory goodness, in order to select the top ranking explanation as *the* explanation.

Lipton explains that explanatory goodness, what he calls 'loveliness' must be distinguished from likeliness, since the aim of IBE is to guide our estimates of likeness on the basis of loveliness. In Lipton's view loveliness is a matter of potential understanding—a lovely explanation is one that would give us a high degree of understanding of the relevant phenomena were it to be true (and, I would add, known to be true).

Two qualifications need to be made concerning the second stage:

(Q1) For the best explanation to be inferred it must be significantly better than its nearest rival. If two competing explanations are both good enough, and one is slightly better than the other, our faith in that slightly better one must be slim. While Lipton does not mention this, it is a clear corollary of his account.

(Q2) For the best explanation to be inferred it should normally, considered on its own, be a *sufficiently good* explanation of *enough* evidence. If our best explanation is a weak explanation even of a large quantity of data (?: 63, 154), or explains only a limited amount of evidence well, then that is some reason to doubt that it is the the actual explanation.

(Later I shall consider amendments to both qualifications.)

Both stages in IBE raise important philosophical questions. A crucial question concerns the first stage. Since it filters out so many logically possible explanations, what confidence can we have that the actual explanation is allowed through? Why

should the imagination of scientists have the capacity to pick on the true explanation among those it creates? The problem here is one that Lipton (?: 152) calls ‘Underconsideration’. The stage 2 ranking is no good at all if the actual explanation hasn’t made it through stage 1 on account of the scientists’ failure to think of it.

Assuming that the actual explanation is among those investigated at stage 2, two problems immediately raise their heads, which Lipton calls ‘Hungerford’s objection’ and ‘Voltaire’s objection’. The former borrows Margaret Hungerford’s line in *Molly Bawn*, that beauty is in the eye of the beholder, to raise the worry that the loveliness of explanations may be too subjective to have any relationship to the truth. Voltaire’s objection suggests that the IBE enthusiast has an unjustified Panglossian faith that the actual world is the loveliest of all possible worlds. Even if loveliness is objective, there will be many worlds where it does not correlate with truth. So why think that truth and loveliness correlate in ours?

In passing I shall mention a hypothesis formulated by David (?), that all these problems have a Kuhnian answer. The fundamental idea is that our standards of goodness are set by Kuhnian exemplars. It is similarity, in the relevant respect, to the paradigms of good science that govern the field in question, that makes for goodness of explanation. That answers Hungerford’s problem. Note that the exemplars are themselves selected on grounds that extend beyond loveliness alone. It is empirical success in solving scientific puzzles where other paradigms have failed that is the principal driver behind the selection of new paradigms. Despite the problems of incommensurability, the development of science is progressive, it is a history of increasing puzzle-solving power. An answer to Voltaire’s objection can build on this, albeit in a non-Kuhnian way. Let’s say for sake of argument that an exemplar has not only puzzle-solving power but also a high truth-content. Then one might expect puzzle-solutions modelled on that exemplar to have at least a better than random chance of latching on to the truth also. The standards of similarity, the qualities that make for explanatory goodness, will then be truth-tropic, even if they are not fully general and sempiternal. Such standards may be local to a particular field at a particular stage in its development, but that does not prevent them from being truth-friendly in their locality. Of course, this depends on starting with an exemplar with high truth-content. But that’s not a problem for two reasons. First, the problem was to show that truth and goodness *could* be correlated, not that they must be. This answer shows how they can be without the world being in any way special. Voltaire’s problem is that we set our standards of loveliness first, and then expect the world to live up to them. This exemplar-based response says that the world itself can play a part in setting the appropriate standards. Secondly, the fact that empirical data and often the puzzles themselves are generated by the world means that as long as there is a genuine puzzle-solving tradition in place, it will have a component that favours truth over falsity; it would not be a surprise that well-established puzzle-solving traditions have exemplars that have high truth-content.

According to this view, explanatory goodness resides in something like Kuhn’s five values—values whose application is determined by exemplars. This differs from Lipton’s conception of loveliness as potential understanding. I don’t intend to adjudicate between these views of explanatory goodness, since I shall argue that in some cases at least we do not need any explanatory goodness at all. That is because, in some cases, inference to the best explanation is inference to the *only* explanation—the problem of Underconsideration notwithstanding.

3 The Semmelweis case (again)

I'll now move on to the first medical case I wish to discuss, the well-known history of Ignaz Semmelweis and puerperal fever. This case is central to Lipton's defence of IBE, having previously been discussed by Hempel and by others. The heuristic advantage is clear: by comparing different accounts of inference and confirmation against a common case, their relative merits can more easily be judged.

There is in Lipton's discussion what seems to me to be an anomaly. Since this is his most detailed case study of IBE, in which various hypotheses are considered that might explain a phenomenon, from which one is selected as being *the* explanation, one might expect some discussion of why the selected explanation is lovelier than the others. We should be told what lovely-making features this explanation has that its rivals lack or possess in lesser degree. But in fact Lipton does not present us with such a discussion. And this suggests to me that the application of IBE in this case does not depend on loveliness or goodness at all.

The Semmelweis case is well-known, and so I shall not spend too much time on the principal facts. In 1844 Ignaz Semmelweis graduated from the Vienna Medical School and decided to study obstetrics. He was appointed assistant to the professor of obstetrics, Johann Klein, first in 1846 and then again in 1847. Klein was responsible for one of the two labour wards at the Allgemeine Krankenhaus, the General Hospital in Vienna. Many poorer women came to the hospital to give birth and of these women a large proportion, up to one sixth in some years, contracted puerperal (or childbed) fever, which was almost always fatal. It was widely known that the death rates were considerably higher in Klein's ward, Division I, than in the other ward, Division II, run by Professor Franz Xavier Bartsch. Semmelweis sought some feature of Division I that would explain its high rate of mortality. These are the principal hypotheses he considered initially:

- (S1) Overcrowding in Division I.
- (S2) Epidemic influences and climate.
- (S3) Rough examinations by the medical students in Division I.
- (S4) Psychological effect of the priest passing through the ward on his way to deliver extreme unction to dying women.
- (S5) Women in Division I delivered on their backs.

In assessing these explanations, we must be careful in deciding what the explanandum is. The explanandum could be, among others:

- (A) The existence of puerperal fever in Division I (and by extension the existence of puerperal fever elsewhere).
- (B) The greater prevalence of puerperal fever in Division I.

The character of the inference is very different depending on which the explanandum is taken to be. In my view it is important to focus on explanandum (B), the difference in rates of puerperal fever and consequent mortality between the wards.¹ The explananda are connected: the principal explanations of the existence of puerperal fever in general might supply explanations of the difference between the two wards; and conversely a successful explanation of that difference might well provide

¹Our explananda concern the existence and rates of puerperal fever, but the data concern rates of mortality from puerperal fever. The mortality rates are good proxies for the morbidity rates since the disease was almost always fatal.

insight into the cause of puerperal fever in general. But these are further inferences, and fraught ones at that, as I shall mention. In the light of this we should consider the explanations (S1)–(S5) as shorthand for explanations of the form ‘X is the cause of the positive difference between the rates of Division I and Division II’, e.g. (S1) should be understood as asserting that overcrowding in Division I is the cause of the greater mortality rate in Division I when compared to Division II.

?: 65–7, 69) noted that hypotheses (S1) and (S2) refer to features that were common to both Division I and Division II. Indeed, because of the desire of expectant mothers to be admitted to Division II rather than Division I, the former was even more crowded than the latter.²

Lipton remarks, however, that the similarity between the wards is nonetheless consistent with one or other of those hypotheses being true. Since no-one thought such factors to be sufficient for puerperal fever, those who maintained such hypotheses would think that they are only *part* of the explanation as to why any particular woman contracted the fever; a full explanation would refer to other factors as well, such as general state of health. Note, though, that this point holds only if the explanandum is (A) rather than (B). But as we have seen and will continue to see, Semmelweis’s principal evidence concerns the *differences* between the two wards. Since Division II had puerperal fever, which could also affect women giving birth at home, Semmelweis was not in a position to directly infer the cause of puerperal fever *tout court*. Naturally, he was indeed interested in the cause of puerperal fever, as the title of his book (?) on the subject demonstrates. But, but as we shall see, the inference from an explanation of (B) to an explanation of (A) makes difficulties for Semmelweis.

According to Lipton’s view of contrastive explanation, to explain the difference between the two wards, we must seek a feature in the history of Division I that is absent from the history of Division II. But hypotheses (S1) and (S2) do not identify such a difference (ignoring the *lesser* degree of crowding in Division I). Therefore they simply cannot be explanations of (B). Those hypotheses, construed as potential explanations of the difference between the two wards, are simply *inconsistent* with the evidence. The same goes for a number of other potential causal factors in a case of puerperal fever that are not mentioned in the list above: inadequate ventilation, excess blood in the circulation, stagnant circulation, disturbances caused by pregnant uterus, decrease in weight caused by emptying of the uterus, protracted labour, wounding of the inner surface of the uterus in delivery, imperfect contractions, faulty involutions of the uterus during maternity, the volume of the secreted milk, and death of the foetus (? : 47).

Hypotheses (S3)–(S5) do mark a difference between the wards, at least at the beginning of Semmelweis’s investigations. (S3), though, was hardly a difference. For as Semmelweis pointed out, the roughness of the handling by the students was negligible compared to the trauma of childbirth itself, and the *difference* in roughness

²In which case, one might ask, why would (S1) even have been raised? The principal answer is that while women were admitted to the two wards on alternating days Sunday through Friday, from Friday to Sunday afternoon, women were admitted to Division I. Furthermore, Division II was instituted in order to relieve overcrowding in Division I. So historically there had been a problem of overcrowding in Division I, until the difference in mortality became widely known. An additional reason is that overcrowding was a widespread problem in European hospitals, with several patients sharing a single bed being a common occurrence. In Vienna, however, one patient per bed was the rule. Nonetheless, the relationship between overcrowding and puerperal fever was a natural one for doctors to consider.

between the students and midwives would have been even smaller in comparison.³ Hence hypothesis (S3) seeks to explain a large difference between the two divisions, the fact that the mortality rate in Division I was three times that in Division II, by appeal to what is at most a tiny marginal difference. Certainly, in some set-ups, incremental changes can have significant effects; but no doctor would suppose *this* to be such a case. In my view such a hypothesis is not merely implausible—Semmelweis could rule it out as inconsistent with what he knew about how trauma affects disease.⁴

With respect to hypotheses (S4) and (S5) Semmelweis pursued the policy of seeking to eliminate the differences between the wards referred to in a given hypothesis. Thus the priest agreed to take a different route, avoiding Division I; and women in that ward delivered on their sides: again in both cases without any diminution in death. Semmelweis was thus able to generate evidence inconsistent with (S4) and (S5) and thereby eliminate them from his enquiries.

At this point, discussions of the Semmelweis case mention the fact that while on holiday in Venice in early 1847, Semmelweis's colleague Jakob Kolletschka died of a wound incurred during a post-mortem examination. In his illness Kolletschka showed the same symptoms and on autopsy the same lesions as found in women who suffered and died from puerperal fever. This led Semmelweis to his final hypothesis, that the parturient women were being infected with cadaveric matter transmitted by medical students from the autopsies that they had been carrying out beforehand. I divide Semmelweis's hypothesis into two components:

(S6a) Women in Division I were infected during examination by medical students. (S6b) The infectious agent was 'cadaveric matter' imported by the students after carrying out autopsies.

Kolletschka's death is often presented as a key piece of evidence, one that (S6) can explain whereas the other hypotheses cannot. Consequently (S6) is, in this respect at least, a better explanation than the others. I believe, however, that the importance of Kolletschka's death lies elsewhere. At this time, the leading explanation offered of puerperal fever, along with many other diseases, was the miasma theory, according to which diseases are often caused by bad airs that are themselves effects of geography and climate, and can be caused by stagnating water, rotting organic material, overcrowding and the like. This is the theory covered by (S2). Note first that in terms of being able to explain other facts, supporters of the miasma theory of disease would argue that their theory explains a huge amount of data, such as the fact that some diseases, such as malaria, are common in low-lying marshy areas, why diseases such as cholera are more prevalent at sea-level than at altitude, why many diseases, such as typhoid and cholera are more prevalent in crowded, unsanitary cities than elsewhere, why improvements in sanitation lead to diminution in

³Lest one should imagine that the midwives were particularly gentle, consider the comment of Semmelweis's colleague Jakob Kolletschka, "It is here no uncommon thing for midwives, especially in the commencement of their practice, to pull off legs and arms of infants, and even to pull away the entire body and leave the head in the uterus. Such occurrences are not altogether uncommon; they often happen." (*Lancet* 2 (1855): 503. Quoted in ? : 126, fn. 5.) In mitigation of the midwives, one should note, as ? do, that many of their patients were women from impoverished backgrounds who had suffered from rickets as children. Rickets can often lead to a malformed pelvis, resulting in difficulties in childbirth when adult.

⁴As it was, Semmelweis sought to minimize the difference by excluding foreign students from Division I, who, he thought, would be the least gentle in their examining. That, of course, had no effect on the rate of infection in that ward.

disease. With respect to the evidence concerning puerperal fever in particular, at the Vienna General Hospital and elsewhere, the miasma theory would explain why puerperal fever comes in epidemic waves and varies seasonally, being worse in winter than in summer. Against this mass of evidence, the fact of Kolletschka's death counts for every little. So if we are considering explanations of (A), then Kolletschka's death counts for very little. But if we are considering (B), Kolletschka's death is evidentially otiose. For the rival (S2), as an explanation of the difference between the two wards, is already refuted by the evidence, as we saw above. And, more importantly, Semmelweis generated *the* crucial piece of evidence when he insisted on the students washing their hands in chlorinated water before examining the women, the mortality rate in Division I fell to equalling that in Division II. The headline figures are these: the percentage mortality rates for the six years 1841–1846 were: Division I—9.92, Division II—3.88, and for the twelve years 1847–1858 were: I—3.57, II—3.05 (figures from ? : 159–81). This crucial fact clinches the argument in favour of (S6a) independently of the evidence concerning Kolletschka.

The significance of Kolletschka's death is that it drew Semmelweis's attention to a difference between the midwives and the students that might otherwise have gone unnoticed—the fact that they attended autopsies and carried out dissections before performing examinations in the maternity wards. While he could not eliminate this difference, since he didn't control the students' timetable, he could isolate it causally, which amounts to the same thing, by insisting on hand-washing.

The evidence concerning Kolletschka enabled Semmelweis to do something else, to formulate a specific hypothesis concerning the infection, that it was due to 'cadaveric matter' being transferred from a dissected body to the uterus of an unfortunate mother via the hands of the students. Above I divided (S6) into a less specific claim (S6a), that some kind of infection from the medical students is responsible; and a more specific claim, that cadaveric matter from autopsies is responsible. This distinction is important because only (S6a) is verified by the evidence. Although (S6b) is rendered plausible, it is far from verified (and indeed it is strictly false). To be precise, the evidence verifies the claim that the explanation of (B) is some property of the hands of the medical students that is removed when they are washed. It strongly supports the claim that this property is related to and a causal consequence of the presence of the students at the autopsies, but without verifying this claim, and it lends some support, but much less, to the claim that the property in question is the presence of cadaveric matter. I should make clear that the notion of 'infection' I am using here is a very weak one, and does not imply any commitment to a modern germ theory. Rather it is intended to capture an idea that would have been familiar to Semmelweis's contemporaries, that of contagion, an idea that goes back to Frascotoro in the seventeenth century. The core of the idea is that diseases can be spread from individual to individual by the conveyance of some material medium between them. Frascotoro hypothesized the medium to be 'seminaria' (seeds), but tells us little about them, which is why I say that the core idea is that there is some material medium of transmission. Semmelweis's insistence on 'cadaveric matter' is a specific version of this theory.

I make these distinctions, even though Semmelweis did not, in order to make two points. First, as I shall go on to explain, making these distinctions will allow me to demonstrate my principal thesis, that the evidence can lead us not simply to the best explanation of the evidence, but also, on occasion, to the *only* explanation of the evidence. (S6a) is a hypothesis of which this is true, but (S6b) is not.

Secondly, I suggest that one reason why Semmelweis failed to get his views accepted is that he did not distinguish between (S6a) and (S6b), and argued strongly in favour of (S6b) which was only partly supported by the evidence. Furthermore, Semmelweis did not clearly distinguish between explananda (A) and (B). Although Semmelweis's most effective evidence concerned the difference between the two wards, his ultimate aim was to explain the causes of puerperal fever *tout court*—all the cases in Division I, and in Division II, and elsewhere. This was because he insisted on a single cause for all cases. But his evidence did not support such a view. For example, it was unclear how cadaveric matter could explain the deaths in Division II and elsewhere. Semmelweis's explanation was that in such cases the women were self-infecting, due to internally decaying matter. Such an explanation seemed *ad hoc*. And while ? (81) noted that street births showed a lower rate of mortality than Division I, he could not explain why home births showed a significantly lower rate of mortality (circa 0.5%) than even Division II (over 3%)—if the cadaveric hypothesis implied that the deaths in Division II were unavoidable self-infections, then one would expect a comparable rate of self-infection among mothers giving birth at home.

Furthermore, the cadaveric hypothesis was not even novel. A commonly held alternative to the miasma theory of puerperal fever was the view that it is caused by internal putrescence, the rotting of the patient's own internal flesh and organs. For example, Dr John Clarke (cited in ? (43) held that tight stays and petticoats and the weight of the baby in the uterus detained faeces in the intestine causing putrescence. Getting people to believe a new theory may be difficult enough, but it is often even more difficult to get them to believe an old theory they regard as having been refuted. The principal piece of evidence against such a theory is the fact that puerperal fever was an epidemic disease which could afflict a population particularly severely for a number of years. Additionally it was seasonal, with winters being particularly bad. ? (122) explained the latter by reference to the greater diligence of the student doctors in winter months, and while that may have been an exacerbating factor, this seasonal variation was not limited to teaching hospitals (which is another reason to focus on explanandum (B) rather than (A)).

To conclude: if we, unlike Semmelweis, restrict our hypothesis to (S6a) and our explanandum to (B), then we see that the evidence forces us to that conclusion by eliminating all potential alternatives. In this case, inference to the best explanation reveals an important kind of limiting case—inference to the only explanation.

In Lipton's model of inference to the best explanation, the loveliness of the hypotheses is central to their epistemic status: the rank order of their epistemic credibility should follow the rank order of their explanatory loveliness. But the epistemology of the aetiology of puerperal fever is not like this. Semmelweis considered six hypotheses, but he did not rank these according to their loveliness. It wasn't that infection via the doctors' and students' hands was a lovelier hypothetical cause of the difference in level of puerperal fever than the presence of a dolorous son of the church. That evidence didn't simply show the priest hypothesis to be unlovely, it showed it to be outright false. Likewise for all the other hypotheses considered by Semmelweis, with the exception of the infection hypothesis (S6a). Thus Semmelweis had no need to consider the loveliness of these hypotheses, and so it is no surprise that Lipton does not discuss their loveliness either.

4 HIV and AIDS

I shall now turn to a more recent case in medical history, the story of the discovery of HIV and the cause of AIDS. The initial phase involved the identification of a syndrome that needed explaining. In June 1981 a report was published concerning the appearance of a rare form of pneumonia, *Pneumocystis carinii* in five homosexual Californian men. *Pneumocystis carinii* had otherwise only been observed in individuals who had undergone medical therapies involving immunosuppression. The following month a second report appeared, discussing the cases of twenty-six young homosexual men with Kaposi's sarcoma, an unusual form of skin cancer, normally found only in men in their 70s and then usually only those of Mediterranean origin. Moreover, four of these had *Pneumocystis* also. Shortly thereafter a further ten cases of *Pneumocystis* were revealed in California. As the Centers for Disease Control (CDC) commented, "The apparent clustering of both *Pneumocystis carinii* and Kaposi's sarcoma among homosexual men suggests a common underlying factor" (? : 14). The clustering of symptoms in a manner indicative of a common cause is a syndrome, in this case initially called GRID, Gay-Related Immune Deficiency Syndrome, and then AIDS, Acquired Immune Deficiency Syndrome.

What explains the existence of this syndrome? What causes AIDS? Researchers considered four hypotheses as follows:

(A1) Recreational drugs. Initially a contaminated batch of 'poppers' (amyl nitrate) was suspected. And then it was considered that excessive use of certain recreational drugs, even if not contaminated, might depress the immune system.

(A2) Some researchers hypothesized that the very high incidence of familiar sexually transmitted diseases among certain sexually very active men might overload the immune system and cause it to fail. This might also explain the appearance of AIDS among intravenous drug users who shared dirty needles—the repeated taxing of the immune system by foreign matter and infections overloads it and makes it unable to fight off opportunistic infection.

(A3) Bacterial infection—infection by a bacterium, probably hitherto unknown.

(A4) Viral infection—infection by a virus, probably hitherto unknown.

To these we may add:

(A0) There is no common cause—the clustering is entirely accidental.

(A0) is the null hypothesis. Abductive inference assumes that there is something in need of explanation. If there is nothing to explain nothing counts as the best explanation of it. Individual events or facts typically do need explanation. If someone falls ill with red pustules over arms, chest, and legs, that needs explanation. As I shall discuss later, that may not be true for all individual events, and certainly not for population level events. For what might appear to be a population level phenomenon of interest may after all be nothing of the sort—just a chance coincidence. Why did I get six sixes in a row? I might have been using a loaded die. But perhaps I was just lucky, which is to say, there is no explanation. Likewise the co-occurrence of certain symptoms in a small number of people might be a coincidence. The fact that the CDC said that the clustering *suggested* a common underlying factor indicates that for them the null hypothesis had not been ruled out. But as numbers

rise, the chances of a coincidence recede rapidly. In Semmelweis's case, the null hypothesis is that there was no medical difference between the wards. By chance the women assigned to Division I were individually more susceptible to puerperal fever than those assigned to Division II. However, Semmelweis's statistics covered so many women and such a continued and dramatic difference between the two wards that the chance of that difference being mere chance was absolutely tiny. Semmelweis's intuition is confirmed not only by common sense, admittedly unreliable as concerns matters of statistics and probability, but also by modern statisticians (cf. ?). Likewise the number of cases of rare symptoms, often overlapping, all related to an impoverished immune system, and in many cases found amongst homosexual men, means that one can conclude in the AIDS case that the null hypothesis is false. There is indeed a genuine syndrome needing explanation.

The key piece of evidence which refuted the lifestyle-related hypotheses (A1) and (A2) was the discovery of AIDS among haemophiliacs. In 1982 several haemophiliacs were found to be suffering from the syndrome, as were a number of people, both men and also women, who had received blood transfusions, including a twenty month old baby. Among the donors of the blood received by the baby was one man who developed AIDS less than a year after donating. While such evidence points to a blood-borne infection, it also serves to exclude the hypotheses (A1) and (A2), since now numerous individuals were beginning to be diagnosed with AIDS who simply did not participate in drug-taking or very active sex. Indeed, this evidence serves to refute pretty well any lifestyle-related hypothesis, since there are no habits shared by the haemophiliacs, the gay men, and the transfusion recipients, that are not shared also by pretty well everyone else.

To my mind, it is difficult to think of any hypothesis compatible with the evidence of the haemophiliacs and transfusion recipients that does not take the cause of AIDS to be an infectious agent. If instead of distinct hypotheses (A3) and (A4) we had a more general hypothesis, that AIDS is caused by an infectious agent, then that hypothesis is confirmed, by refuting the null hypothesis and all other hypotheses inconsistent with this one. Having established that AIDS is an infectious disease, the next task is to identify the kind of infection. The two obvious candidates are bacterial and viral. The evidence already obtained rules out bacterial infection. This is because the blood product used by haemophiliacs, the clotting agent factor VIII, is obtained from donated blood by a process that involves, among other things, filtration. Filtration removes bacteria, and so the bacterial hypothesis can be excluded.

With the bacterial hypothesis refuted, it is natural to turn to the viral hypothesis. However, one might wonder whether some other infectious agent could be responsible: not every infection is bacterial or viral; the other possibilities include fungi, protozoa, and multicellular parasites. In fact filtration removes all of these agents also. Arguably it is conceivable that some hitherto undiscovered kind of filterable agent could be responsible. We now know that there are such agents, although most are virus-like, such as satellite viruses and viroids, and typically these require the presence of a true virus, a helper virus, to replicate. However, the first research into prions was being carried out at about the same time as the cause of AIDS was being investigated, and so such a possible cause would not have been considered. Like a virus, a prion, being simply a protein, is filterable. It remains contentious, however, that prion-related disease is caused by a protein-only agent rather than by protein-plus-virus or some other mechanism. Indeed, one of the controversial features of the prion hypothesis, is that it appears to be inconsistent with the central dogma of molecular biology. The latter says that information can be passed only from nucleic

acid (DNA, RNA) to nucleic acid or to protein, but never from protein to protein or from protein to nucleic acid. But prions are proteins, and so if prions are both the causes and effects of prion-related disease (such as CJD), then there is information transfer from protein to protein. If this objection from the central dogma holds good, then it looks as if *only* a virus, or virus-like organism (satellite virus or viroid) could be the AIDS vector, since such a vector must contain DNA or RNA, but to be filterable it must be smaller than cellular. Being non-cellular, the agent cannot carry the means of its own replication, but must depend upon some external mechanism. That is tantamount to a definition of a virus. Nonetheless, I do not think that such an argument suffices to give us knowledge that AIDS is caused by a virus, since the epistemic status of the central dogma is not sufficiently well established that it amounts to knowledge. Crick's point in calling the claim a dogma was that he felt that despite its importance it was not well supported by the evidence.⁵ Correspondingly it would not be safe to conclude that only a virus is consistent with the evidence mentioned so far regarding the cause of AIDS. What did establish that a virus causes AIDS was the isolation of a particular virus, LAV (lymphadenopathy AIDS-associated virus), by Luc Montagnier in 1983, renamed HIV three years later. In due course HIV was shown to satisfy Koch's postulates with respect to AIDS (Koch's postulates being principles used to establish that a certain infectious agent is the cause of a given disease).⁶

What the AIDS case shows is, again, that the identification of the correct explanatory hypothesis proceeds by the refutation of principal rivals. While the methodology may have a Popperian flavour, the epistemology does not. For the process of elimination raises the likeliness of the remaining hypotheses. The epidemiological evidence ruled out hypotheses such as the overloading of the immune system by drugs or commonplace STDs; indeed it refuted any hypothesis other than those permitting blood-borne infection. That raised the probability that the cause of AIDS is a bacterium or virus. The fact that the infectious agent is filterable rules out bacterial infection, raising the probability further that the cause is a virus. That does not establish the viral hypothesis with certainty, since that evidence is consistent with subviral agents as a cause. Nonetheless, the fact that such agents are rare, means that the virus hypothesis had a high probability, which encouraged Montagnier to search for a virus directly.

5 Eliminative abduction and Holmesian inference

Above I have argued that a key part of understanding Semmelweis's reasoning must be the fact that he *refuted* the competing hypotheses. No doubt one large part of the popularity of Popper's philosophy among scientists is the fact that they recognize the role that refutation does play in science. But the Semmelweis case also shows that the sceptical side of Popper's philosophy, usually ignored by scientists, is unwarranted, at least as a description of scientific practice. For the verdict of subsequent scientists is that Semmelweis did have very good reason for believing his conclusions, at least when framed in a suitably circumspect manner; indeed his ev-

⁵Note however, that this, Crick's informational version of the central dogma from 1958, is not refuted by the many criticisms directed at Watson's pathway account of 1965, popular in textbooks, according to which DNA generates RNA (or more DNA) and RNA generates proteins.

⁶That HIV does satisfy Koch's postulates with respect to AIDS was disputed in some quarters, most notably by Peter ?). But there is now little mainstream doubt on this point.

idence allowed him to *know* that the cause of the differential mortality rate was as stated in hypothesis (S6a).

If that is right, then Semmelweis's reasoning bears a close relation to what I have called 'Holmesian inference' (?), in recognition of the famous dictum, "Eliminate the impossible, and whatever remains, however improbable, must be the truth." (? : 94, 118). (For eliminative induction—and Sherlock Holmes—see also ????) Spelt out more systematically, Holmesian inference has the following structure:

- (i) the fact e_s has an explanation (Determinism);
- (ii) h_1, \dots, h_n are the only hypotheses that could explain e_s (Selection);
- (iii) h_1, \dots, h_{n-1} have been falsified by the evidence (Falsification);
- therefore*
- (iv) h_n explains e_s .⁷

Holmesian inference is clearly deductive, which is why Conan Doyle correctly describes his hero as a master of *deduction*. Therefore in considering whether Holmesian inference can ever lead to knowledge, we need to know whether we can ever be in a position to know that the premises of a Holmesian argument are ever true.

Falsification ought not present a problem. Aficionados of the Duhem-Quine thesis might have their doubts, but I regard these as exaggerated, especially in the hands of Quine. I shall not pursue that more general issue here. To give just one example, it seems clear that the hypothesis that the priest's presence is a cause of the higher mortality rate is simply refuted by the fact that absence is not marked by any reduction in mortality.⁸

Determinism is the denial of the null hypothesis, which states that there is no explanation of the phenomenon in question. For individual macroscopic events, the null hypothesis will usually be false and so Determinism will be true, and knowable. However there are cases where Determinism might fail. It may fail for individual atomic or subatomic events; not all such events have an explanation. For example the decaying of a fissile nucleus does not itself bear an explanation, since that is an intrinsically indeterministic occurrence; it just happens. (Nevertheless, we can explain related facts, e.g. that it was possible for it to decay, or that its chance of decay was p , etc.). When we move to statistical phenomena, we may find that there are borderline cases. The proportion of the population at large which is left-handed is about 12%. Let us imagine that a survey of a lecture audience showed that its proportion of left-handers is 16%. It is not immediately obvious whether this fact has an explanation. This difference from the national average *could* be just a statistical fluctuation. Classical significance testing aims to quantify this, by telling us what the chance would be of a group of just this size having a proportion equal to 16% when it is chosen from the population at large in a manner independent of any possible causal factor (i.e. 'randomly'). The null hypothesis is the hypothesis that the group can be regarded as chosen independently of left-handedness.

⁷For discussions of Determinism and Selection see ? : 131).

⁸I do note, however, Alex Broadbent's (?) response on behalf of the Liptonian model, that the removal of the priest without change in mortality does not refute the priest hypothesis but just makes it incredibly unlovely: "... it seems to me that Semmelweis did not refute the priest-hypothesis when the priest was removed. Maybe the effects of religion are delayed, and Semmelweis did not wait long enough for the level in the two wards to equalize. Maybe there was some confounding factor: maybe the sudden absence of the priest caused concern among the inmates, as alarming as they had found his presence. Hypotheses of this sort can be devised that are consistent with the claim that the priest's route caused the difference between the two wards. But they are incredibly unlovely"

Selection is more controversial. It is often contended that for any set of evidence there is an unlimited number of hypotheses consistent with that evidence. Note that Selection requires the hypotheses not merely to be consistent with the evidence but also to explain the crucial phenomenon in question. And one ought to construe ‘explanation’ in a reasonably robust way (as does Lipton). Merely deducing e from some proposition h plus certain conditions c does not amount to an explanation of e —for example when $h = \neg c \vee e$. Still, a sceptic may still insist that any phenomenon may have an unending range of mutually inconsistent potential explanations. But for Falsification to retain its plausibility, the range of explanations needing to be considered must be finite, indeed finite and reasonably small.

In some circumstances one can engineer an experimental setup so that all potential explanations are excluded bar one—we get to know Falsification and Selection simultaneously. Much laboratory science is like this: one varies one factor at a time and so one is able to infer from an observed difference that it is caused by that single factor. This is Mill’s method of difference. Much medical science is based on the method of difference. The Randomized Controlled Trial is intended to be the method of difference writ large. If there is a sufficiently large difference in outcome in a sufficiently large trial so that the null hypothesis can be eliminated (i.e. Determinism holds in this case), then one can infer that the treatment is the only possible explanation of that difference and hence is *the* explanation of that difference. Sometimes at least, we can be in a position to know that there is only one explanation of the evidence. Arguably Semmelweis ended up in something like this position.

In many cases we will not be able to engineer a position in which Mill’s method applies. The problem is that there may be potential explanations that are sufficiently out of the ordinary that they never get considered, but which are not refuted by our evidence. This is just the problem of Underconsideration mentioned already. And insofar as it is a problem, it is one that afflicts not only Holmesian inference but also Lipton’s model of IBE. Whether one chooses a favoured explanation at stage 2 by ranking or by refutation of competitors, the choice will fail to be known to be the explanation, if one has failed to consider the actual explanation or indeed false potential explanations that ought to have been considered.

The answer to the problem of Underconsideration is to appeal to general externalist epistemology.⁹ In order to know that an instance of Selection is true it is not required that one have considered *every possible* hypothesis consistent with one’s initial evidence. It will normally be sufficient that one has considered all the potential explanations that are true in nearby possible worlds. The sense of ‘could’ in Selection is not the philosopher’s liberal one meaning ‘in some possible world’ but a more restrictive one, exemplified by the true statement, as uttered just before the 2008 US election, ‘either McCain or Obama could win the election on Tuesday, but Ralph Nader could not’. For example, one may take a central feature of knowledge to be the fact that it is closely related to *safe* believing. One believes p safely if p is true in all conditions that are in fact similar to the actual condition—in terms of possible worlds, S believes p safely if p is true in all nearby possible worlds. While it is too simplistic to say that safe believing suffices for knowing, we may nonetheless employ the general idea that to know that p does not require that one’s evidence rule out *all* possibilities, however remote. The issue of underconsideration is not problematic so long as all the hypotheses that are not considered, even if they are

⁹David ? makes a related appeal to externalist/naturalized epistemology, but instead of requiring Selection to be known, he regards it as sufficient that one has a reliable disposition to infer (iv) from (iii) alone. For my response see ? : 15.

consistent with our evidence, are true only in remote possible worlds. If that condition is met Selection comes out as true and knowable.

Having articulated Holmesian inference, I now turn to the relationship between that inference pattern and the two cases we have examined above. Determinism holds in Semmelweis's case because the statistical difference between the wards was large enough to be clearly no accident. Likewise, when the number of cases of GRIDS/AIDS, a distinctive and hitherto unusual combination of symptoms, was sufficiently high, researchers could know that they had a new disease on their hands with a specific cause—the null hypothesis, (A0) can be dismissed. Falsification holds with respect to the set of hypotheses (S1)–(S5) discussed: Semmelweis refuted hypotheses by gathering evidence inconsistent with them. Likewise, the evidence that AIDS researchers possessed allowed them to refute hypotheses (A1)–(A3).

Does Selection hold for our two cases? It is not obvious that the six hypotheses considered by Semmelweis are all the potential explanations there could be, even that they include all the hypotheses that could be true in nearby possible worlds. Above I listed a range of hypothetical causes of puerperal fever that Semmelweis's contemporaries proposed and which might have marked a difference between the two wards. Consequently, the elimination of the other hypotheses did not itself establish the truth of the infection hypothesis (S6a). Nonetheless, the method employed by Semmelweis to eliminate specific hypotheses could be used to eliminate all hypotheses except his favoured one. If the two wards are made to be identical in every causally relevant respect except one, and the difference in mortality remains as before, then the remaining respect must be the cause. As it happens, the elimination of (S1)–(S5) made (S6a) highly likely and allowed Semmelweis to test it by direct intervention. Likewise, as explained, the elimination of various hypothetical causes of AIDS made the viral hypothesis very likely, but did not directly establish it. But by making the viral hypothesis likely it made the difficult investigation of a viral hypothesis epistemically more profitable.

Does this suggest that the Holmesian model in fact does not, strictly speaking, ever apply, because we may always reasonably doubt Selection to be true? I don't believe so. Some scientific investigations are too complex for them to be summarized as exemplifying any single argument form. Ernst Mayr characterized Darwin's origin of species as 'one long argument', where this one long argument is made up of many arguments some of which may individually lend only partial support to their conclusions. Semmelweis could not use Selection to come to know the truth of (S6a); nonetheless the elimination of rival hypotheses made (S6a) likely. This led to Semmelweis being able to use Mill's Method of Difference, which is itself a special case of Holmesian inference, to come to know that (S6a) is true, not merely that it is likely. Something similar may be said in the identification of HIV as the cause of AIDS, since there again, I conjecture, the use of Koch's postulates can be considered a specific case of Holmesian inference.

6 Eliminative abduction and inference to the best explanation

If Holmesian inference depends on the elimination of hypotheses by refutation, does explanatory power play no epistemic role? My case for Holmesian inference notwithstanding, I agree with most of what Lipton says about the confirmational

power of inference to the best explanation, even when the best explanation is not the only one consistent with the total evidence. In discussing the epistemology of some reasoning process one must distinguish what is sufficient to produce knowledge from what is sufficient to produce good reasons to believe. Philosophy of science has tended to ignore knowledge as a category and focus solely on degrees of confirmation. One reason for this is a residual scepticism that affects even many scientific realists, that we never get to know any theoretical truths since none of them are strictly true—a view I think is badly mistaken. Epistemologists outside the philosophy of science do make the kind of distinction I have made. Thus it is commonly thought that one cannot know that one's lottery ticket will lose, but the evidence that it is only one ticket in a thousand gives one a good reason to believe that it will lose. Since I don't think that even statistical cases can be treated like a lottery, I shall not discuss this analogy. Its purpose is simply to point out that a discussion that tells us a great deal about when our beliefs are well-supported by the evidence may well leave unanswered important questions about when our beliefs amount to knowledge. In my view Lipton's IBE is our best account of confirmation in theory choice, but needs to be supplemented by Holmesian inference if we are to understand when we can gain *knowledge* of the truth of an explanatory hypothesis.

Clearly there must be a connection between the correct account of confirmation and the correct account of conditions for knowledge. A key condition is that the conditions for knowledge, in the case of knowledge gained by inference from evidence (as opposed to non-inferential knowledge, such as perception), are such that knowing entails a high degree of confirmation. How do IBE and Holmesian inference relate? The answer is clear: while IBE sorts hypotheses according to their degree of ability to explain the evidence, Holmesian inference corresponds to the case where the sorting is extreme: the best explanation comes to the top because all its competitors are rejected as explanations of the evidence since they are inconsistent with it. The only explanation is the best by default. Lipton proposes that IBE should be regarded as a surrogate or heuristic stand in for Bayesian conditionalization. So, for example, our estimate of the likelihood ratio, $P(e/h)$ is often given by considering the potential explanatory relationship between h and e . Now consider a Holmesian case where all explanatory hypotheses have been falsified bar one. In such circumstances the falsified hypotheses have zero loveliness and zero posterior probability.¹⁰ And the remaining hypothesis is now the only hypothesis that has any loveliness. Its posterior probability must be one, since the posterior probabilities of all its competitors sum to zero. In such a case our original qualifications, (Q1) and (Q2), no longer apply. To be worth inferring a hypothesis need not meet a minimum threshold of loveliness (other than zero loveliness) in every case. In the Holmesian case it can have quite a low loveliness and still be inferable, since its posterior probability is one. The point of Holmes's dictum ('Eliminate the impossible, and whatever remains, however improbable, must be the truth.') is that a low prior and low loveliness may be translated into a posterior of 1, without increasing loveliness at all, so long as all competitors are refuted.

¹⁰Cannot a hypothesis be lovely, even if known to be false, such as Newtonian mechanics? Some aspects of loveliness are independent of the evidence, such as simplicity. But other aspects are evidence relative, such as whether they provide a unified explanation of the evidence. And it is difficult to see how the latter can ignore the incompatibility of hypothesis and evidence. If, however, one does have a conception of loveliness that is compatible with known falsity, then Lipton's model of IBE will need a further intermediate stage where plausible and lovely but falsified hypotheses are filtered out.

7 Conclusion

Lipton's account of IBE is still our best theory of how scientists distribute their credences among theories on the basis of their explanatory power. Taking our cue from his discussion of the Semmelweis case, where loveliness does not play a role in theory choice, we can identify a special case of IBE where the evidence leaves us with only one hypothesis from among those generated at Stage 1 of IBE. Inference to the only explanation is a limiting case of inference to the best explanation.

If my discussion of Holmesian inference is correct, we can see how in this limiting case at least IBE can lead to knowledge: we have a simple deductive inference from known premises. This raises the question, whether IBE can lead to knowledge in cases that fall short of Holmesian inference. That is a question for another time. But it seems to me that in a case where we can only rank because not all considered hypotheses have been refuted, then we cannot know that the loveliest hypothesis is true, since we do not yet know that an unlovely but unrefuted competitor is false. If such considerations are correct, then Holmesian cases are not simply special cases of IBE, they are *very* special cases, since they are precisely the cases where IBE leads to knowledge.

8 Acknowledgments

As is exemplified by this paper, my own thinking about problems in the epistemology of science is not simply deeply indebted to Peter Lipton, but has developed within a framework that Peter constructed. Peter's work and the way in which he wrote philosophy and indeed the way in which he approached the subject, will remain exemplary for me and for many others for a long time to come. Less apparent to others is the way in which Peter promoted the ideas and careers of younger philosophers. I, with many others, benefitted hugely from Peter's kindness, his advice, and his unending willingness to engage in philosophical discussion. We have so much to be grateful to him for.

As regards this paper, I am grateful for very helpful comments to audiences at the Universities of Bristol, Edinburgh, Exeter, Hertfordshire, and Stirling, and at the Peter Lipton memorial conference held in Cambridge on 1 November 2008. I am grateful also to Alex Broadbent, Mark Sprevak, and Anjan Chakravartty for stimulating conversations on the topics raised in this paper, and again to Anjan Chakravartty for comments on a draft.

References

- Bird, A. 2005. Abductive knowledge and Holmesian inference. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, pp. 1–31. Oxford: Oxford University Press.
- Bird, A. 2007. Inference to the only explanation. *Philosophy and Phenomenological Research* 74: 424–32.

- Broadbent, A. 2008. Holmesian elimination and Liptonian loveliness. *Peter Lipton Memorial Conference, 1 November 2008*.
- Buyse, M. 1997. Opening address: A statistical tribute to Ignaz Philip Semmelweis. *Statistics in Medicine* **16**: 2767–72.
- Carter, K. C. and B. R. Carter 2005. *Childbed Fever: A Scientific Biography of Ignaz Semmelweis*. Edison, NJ: Aldine Transaction.
- Conan Doyle, A. 1953. The sign of four. In *The Complete Sherlock Holmes*, Volume I. New York: Doubleday.
- Connor, S. and S. Kingman 1989. *The Search for the Virus*. Harmondsworth: Penguin Books.
- Duesberg, P. 1996. *Inventing the AIDS Virus*. Washington D.C.: Regnery.
- Earman, J. 1992. *Bayes or Bust?* Cambridge MA: Bradford.
- Gillies, D. 2005. Hempelian and Kuhnian approaches in the philosophy of medicine: the Semmelweis case. *Studies in History and Philosophy of Biological and Biomedical Sciences* **36**: 159–181.
- Kitcher, P. 1995. *The Advancement of Science*. New York: Oxford University Press.
- Lipton, P. 2004. *Inference to the Best Explanation* (2nd ed.). London: Routledge.
- Nuland, S. B. 2003. *The Doctors' Plague*. New York: Norton.
- Papineau, D. 1993. *Introduction to Philosophical Naturalism*. Oxford: Blackwell.
- Semmelweis, I. 1983. *The Etiology, Concept, and Prophylaxis of Childbed Fever* (K. Codell Carter trans and ed.). Madison, WI: University of Wisconsin Press.
- von Wright, G. 1951. *A Treatise on Induction and Probability*. London: Routledge and Kegan Paul.
- Walker, D. 2009. *A Kuhnian Defence of Inference to the Best Explanation*. Ph. D. thesis, University of Bristol.