VILNIUS UNIVERSITY

Adolfas Mackonis

INFERENCE TO THE BEST EXPLANATION: THE LIMITS OF TRUTH-CONDUCIVENESS

Doctoral dissertation
Humanities, Philosophy (01 H)

Vilnius, 2011

The dissertation was prepared at Vilnius University during 2007-2011

Research supervisor:

Prof. Dr. Marius Povilas Šaulauskas (Vilnius University, Humanities, Philosophy – 01 H)

Counsellor:

Assoc. Prof. Dr. Nijolė Radavičienė (Vilnius University, Humanities, Philosophy – 01 H)

VILNIAUS UNIVERSITETAS

Adolfas Mackonis

GERIAUSIO PAAIŠKINIMO IŠVEDIMO PATIKIMUMO RIBOS

Daktaro disertacija
Humanitariniai mokslai, filosofija (01 H)

Vilnius, 2011

Disertacija rengta 2007-2011 metais Vilniaus universitete

Mokslinis vadovas:

prof. dr. (hp) Marius Povilas Šaulauskas (Vilniaus universitetas, humanitariniai mokslai, filosofija – 01 H)

Konsultantė:

doc. dr. Nijolė Radavičienė (Vilniaus universitetas, humanitariniai mokslai, filosofija – 01 H)

Contents

Introduction

Relevance of the thesis. The thesis explicates and analyses the inference to the best explanation (IBE): a meta-theoretical principle of inference of scientific propositions. Scientists apply IBE both in the context of discovery and in the context of justification of scientific theories. More particularly, scientists seek such theories that would explain observable phenomena, cohere with already accepted scientific knowledge, be simple, or unify explanations of different kinds of phenomena, and scientists argue for the acceptance of their theories on the grounds that they do indeed explain observable phenomena, cohere with already accepted scientific knowledge, are simple, or unify explanations of different kinds of phenomena. *Any* argument for the truth or reality of a theoretical term, concept, entity or theory in general is an instance of IBE. IBE is a fundamental component of theoretical reasoning in general and of scientific practice in particular.

IBE is not bound to the context of science. It is also prevalent in medical diagnosis, criminal investigation, judicial argumentation, technical troubleshooting, common sense reasoning and other contexts where one wants to find out the causes of some phenomena. For example, IBE is so prevalent in medical diagnosis or criminal investigation that it even receives a substantial amount of attention in fictional representations of these practices, e.g., the differential diagnoses in the TV-series "House" or the deduction method used

by Sherlock Holmes in Arthur Conan Doyle's stories.

Because of its role in the scientific and commonsense reasoning contexts, IBE is also an important topic for philosophical research in current analytical philosophy of science and epistemology. The truth-conduciveness of IBE is of ultimate importance for the problem of scientific realism. According to scientific realism (e.g., Kitcher 1993; Leplin 1997; Niiniluoto 1999a; Psillos 1999), the best explanation for the empirical success of mature scientific theories is their truth. Scientific theories are the products of IBE. Therefore, if scientific realism is true, then IBE seems to be truth-conducive. On the other hand, IBE deprived of truth-conduciveness makes scientific realism less viable.

IBE is often considered as the development of C. S. Peirce's idea of abduction (Peirce 1932-1958). Nevertheless, the term inference to the best explanation was introduced only in 1965 by Harman (1965). The most vigorous defense of the truth-conduciveness of IBE is presented by Lipton (1993; 2001a; 2004) and Psillos (2002; 2007; 2009d). Different arguments for the truth-conduciveness of IBE are also put forward, for example, by Carruthers (1992), Goldman (1990), Josephson and Josephson (2003), Glass (2010), Niiniluoto (1999b) and Thagard (2007b). Van Fraassen (1980; 1989), on the other hand, presented the most important arguments against IBE.

The aim of the thesis. The thesis is going to analyze the problem of truth-conduciveness of IBE: does the nature of IBE warrant the truth-conduciveness of IBE? The problem has two aspects: given a pool of potential explanations, which one of the potential explanations is the best, and, given that we identified the best explanation, how confident can we be that it is true. More particularly, firstly, the thesis explicates the concept of IBE, i.e., it analyzes what the

concept of explanatory power stands for: how to obtain a pool of potential explanation; given a pool of potential explanations, which one of these is the best; and is it always possible to discriminate the best explanation? Secondly, the thesis evaluates the tenability of the truth aspirations of IBE, i.e., it analyzes to what extent a hypothesis can be claimed to be true, given that it is the best explanation for some phenomenon of interest?

Claims of the thesis. The thesis claims that even though IBE could facilitate the determination of probability distributions and is a wide psychological practice, due to the bad lot argument, possible incommensurability of explanatory virtues, pessimistic induction and better-safe-than-sorry argument, all the four ways of justifying IBE in terms of truth-conduciveness cannot be taken for granted which leaves IBE only accidentally valid. More particularly, the thesis argues for the following claims:

1. IBE is a form of material inference that ascribes truth to the hypothesis that has the highest degree of explanatory virtues among its competitors: it is the most consistent with approved background knowledge, unifies the most the relevant phenomena, is the deepest explanation, and is the simplest explanation. This explanationist account is more fundamental than probabilistic accounts of IBE, but coherence should not be treated as one of the explanatory virtues.

2. Currently, there are four basic ways of justifying the truth-conduciveness of IBE that can be discerned in literature: reliabilist-coherentist, evolutionary, probabilistic and empirical-historical.

3. None of the discerned ways of epistemic justification of IBE grants the truth-conduciveness of IBE. They are undermined by the bad lot argument, the argument of pessimistic induc-

tion, the better-safe-than-sorry argument, and the possibility of contradicting orders of explanatory power. Therefore, being the best explanation does not grant truth.

4. IBE is warranted pragmatically, i.e. IBE is a widespread psychological practice, there is no better method of ampliative inference and its use helps to successfully cope with the world.

Methodology of the thesis. The ongoing discussion on IBE is meaningful only if the correspondence theory of truth is held to be correct. Thus the thesis is going to assume the correspondence theory of truth. The thesis also assumes the naturalized epistemology perspective, in which findings from the sciences that study human reasoning are brought to bear on questions in philosophy. Even though, deprived of truth-conduciveness IBE makes scientific realism less viable, this would not refute scientific realism: scientific realism can be true even if IBE is not truth-conducive. Therefore, the thesis would not argue for nor against the truth of scientific realism. A clear distinction is drawn between IBE and abduction. Nevertheless, claims about abduction are used to illustrate points about IBE: this happens only if the author of the claim does not discriminate between IBE and abduction and means the same thing by the both concepts.

One difficulty in discussing inference to the best explanation lies in the ambiguity of the term denoting it. IBE can denote a particular kind of derivation. IBE can denote the conclusion of a derivation. Finally, IBE can also denote a theory or a set of claims about the properties of inference to the best explanation. In order to avoid this ambiguity the following convention is going to be used in this thesis. "IBE" will stand for IBE as a derivation or process of inference. "A conclusion of IBE" will stand for the conclusion of IBE as a process

of inference. "A theory of IBE" will stand for a theory or a set of claims about the properties of IBE. The term "truth-conduciveness" will stand for the reliability of the method of inference: hence, IBE is truth-conducive if and only if IBE tends to generate true or approximately true conclusions. The term "truthlikeness" will stand for closeness to truth or the degree of verisimilitude: hence, if IBE is truth-conducive, it will tend to generate truthlike conclusions.

Novelty of the thesis. The thesis reconstructs the concept of IBE and examines the results of the reconstruction for internal consistency, conceptual coherence and fit with empirical and historical data. The novelty of this thesis lies in two main aspects. Firstly, the thesis presents an original account of IBE and argues how to evade possible inconsistencies that are present in theories of IBE. It adopts and argues for the explanationist (e.g., Lipton 2004; Psillos 2002) rather than probabilistic (e.g., Glass 2007; Schupbach and Sprenger 2011) approach to IBE. Secondly, the thesis discerns four ways in which philosophers argue for the truth-conduciveness of IBE and shows that these ways are insufficient to grant the truth-conduciveness of IBE. Van Fraassen's (1989) bad lot argument and Stich's (1990) argument about the better-safe-than-sorry beliefs are defended and elaborated to argue against the truth-conduciveness of IBE. Laudan's (1981) pessimistic induction argument is applied to the issue of IBE. An argument about the possible incommensurability of explanatory virtues is produced for the problem at hand.

Structure of the thesis. The thesis has two main chapters that correspond to the two aspects of the main problem delineated above. The first chapter reconstructs IBE and argues for the explanationist account as an adequate account of IBE. IBE is commonly described as consisting of two steps or two filters: the first gathers a pool of potential explanations and the second selects the best explanation out

of the pool of potential ones. Therefore section 1.2 provides an account of abduction as the first step of IBE and the next two sections present two alternative accounts of the second step: the explanationist account (1.3.1) and its defense (1.4.2), according to which, the explanatory power of a hypothesis is the degree of explanatory virtues present in the hypothesis and the probabilistic (Bayesian) account (1.4.1) and its critique (1.4.2), according to which, the explanatory power of a hypothesis is a particular product of the prior probability and likelihood of the hypothesis. The second chapter evaluates whether IBE can be granted truth-conduciveness. The chapter enumerates four basic ways in which the truth-conduciveness of IBE is justified, argues that these means of justification do not grant the truth-conduciveness of IBE and finishes with the section (2.6) which summarizes arguments that refute the truth-conduciveness of IBE. Subsections on psychological adequacy (2.3.1) and ontological commitments of IBE (2.5.1) are apposite to some of the justifications of IBE and are provided in the relevant parts of the second chapter.

Chapter 1

What is IBE?

1.1 Genesis of IBE

IBE is claimed to provide true propositions and beliefs. Despite its claimed importance and high aspirations, IBE is still not a thoroughly investigated and well defined form of inference. According to Lipton, the theory of IBE is "more a slogan than an articulated philosophical theory" (Lipton 2004: 2, 57). More particularly, according to Glass, "IBE faces two key challenges. First, how exactly is IBE to be understood and made precise? [...] Second, what is the connection between explanation and truth? Is there any reason for thinking that the best explanation is likely to be true? Or to put it another way, does IBE track truth" (Glass 2010)? Douven also states that "the exact form as well as the normative status of abduction are still matters of controversy" (Douven 2011).

The term "inference to the best explanation" was coined by Gilbert Harman in 1965 in an article by the same title: "Inference to the Best Explanation" (Harman 1965). Harman argued in the article that there are many warranted non-deductive inferences that are not instances of enumerative induction, therefore IBE and not enumerative induction should be considered to be the main kind of non-deductive inference. The biggest impact was made by Harman's idea

that the explanatory power of a hypothesis determines its truth. In other words, Harman defined IBE as a form of inference, according to which, from a premise that a given hypothesis provides the best explanation among its competitors, one concludes that the hypothesis is true. Harman's work initiated the discussion, but did not provide any argument in support of the truth-conduciveness of IBE. The most vigorous defense of the truth-conduciveness and warrantedness of IBE is found in the works of Peter Lipton (1993; 2001a; 2004) and Stathis Psillos (2002; 2007; 2009d). Truth-conduciveness of explanatory considerations is also defended by Josephson (2001), Josephson and Josephson (2003), Lycan (1988), Thagard (2007b) and Tuomela (1985). Carruthers (1992) and Goldman (1990) argue for the truth-conduciveness of IBE on evolutionary grounds. Because of the acclaimed importance of explanatory considerations, proponents of the truth-conduciveness of IBE are often called explanationists.

There are also more moderate interpretations of IBE. Day and Kincaid (1994) and Ben-Menahem (1990) argue for the truth-conduciveness of IBE as a general inferential strategy rather than a particular form of inference. Niiniluoto (1999b; 2004) argues that IBE combined with empirical testing is the best method of seeking informative truths in science. Kuipers (2004) argues that IBE should infer comparative (best is closest to the truth) rather than absolute (best is true) conclusions.

The works of Bas van Fraassen (1980; 1989) present the most vigorous critique of IBE. Van Fraassen claims that IBE not only does not take us towards truth, but also makes us incoherent. Psillos (1996; 1999) defends IBE from van Fraassen's arguments, but Ladyman et al. (1997) criticize Psillos' defense. Douven (1999; 2002) defends IBE from objections, but does not argue for the truth-con-

duciveness of it. Barnes (1995), Bartelborth (2005), Salmon (2001a) and Schurz (2008) are rather skeptical about the truth-conduciveness of IBE.

The exact account of IBE or of explanatory power is also a debatable topic. Thagard (1978; 1993) and Ylikoski and Kuorikoski (2010) discuss criteria for evaluating explanations. Glass (2007; 2010), McGrew (2003) and Schupbach and Sprenger (2011) propose probabilistic accounts of explanatory power.

Bayesianism is the dominant theoretical approach in the contemporary analytical philosophy of science, epistemology and decision theory. Given that, a reconciliation of IBE with Bayesianism is a lively topic related to an assessment of the truth-conduciveness and general usefulness of IBE. Lipton (2001a), Niiniluoto (1999b), Okasha (2000), Salmon (2001b) and Weisberg (2009) discuss the prospects of reconciling IBE with Bayesianism. On the other hand, Psillos (2004; 2009b) is skeptical about the need for reconciliation.

On a closely related topic of abduction, initiated by Charles Sanders Peirce, the most important studies are written by Aliseda (2006), Gabbay and Woods (2005), Paavola (2006b) and Schurz (2008). The thesis considers IBE to be conceptually different from abduction. Nevertheless, the two concepts are very closely interrelated. For example some researchers do not discriminate conceptually between IBE and abduction, or they use the term "abduction" to stand for IBE (Barnes 1995; Carruthers 2006; Douven 2011; Fodor 2000; Josephson and Josephson 2003; Niiniluoto 1999b; 2004; Plečkaitis 2006; Psillos 2000; 2002; 2004). Some argue that IBE and abduction are conceptually distinct (Campos 2009; Minnameier 2004; Hintikka 1998; McKaughan 2008), and some consider that IBE and abduction overlap to some degree (Aliseda 2006; Gabbay and Woods 2005; Kuipers 2004; Magnani 2001; Paavola 2006a; Schurz

2008; Thagard 2007b).

1.2 Abductive Mechanism of IBE

The concept of IBE is most commonly associated with the concept of abduction. The term "abduction", at least in the history of philosophy, is associated with the name of Charles S. Peirce. Peirce provided a standard definition of abduction:

> The surprising fact, C, is observed;
> But if A were true, C would be a matter of course,
> Hence, there is reason to suspect that A is true. (Peirce 1934: 5.189)

IBE differs from abduction in two fundamental aspects. Firstly, IBE consist of two abduction premises together with an additional third premise, which states that there is no better available explanation than the hypothesis analyzed. Secondly, the addition of the third premise is claimed to grant that the conclusion of the inference is actually true rather than merely possible. Thus the form of IBE is this (Josephson and Josephson 2003: 5; Lycan (1988: 129) and (Psillos 2002: 614) provide similar definitions):

> The surprising fact, C, is observed;
> But if A were true, C would be a matter of course,
> No available competing hypothesis can explain C as well as A does.
> Hence, A is true.

In other words, the conclusion of abduction is only a (logical) possibility: "merely suggests that something may be" (Peirce 1934: 5.171); is "nothing but guessing" (Peirce 1958: 7.219); is "ignorance-preserving" (Gabbay and Woods 2005: 43); "might be defeated"

(Aliseda 2006: 31). The conclusion of IBE "is true" (Harman 1965: 89), "[probably] is true" (Lycan 1988: 129), "is true, or at least approximately true" (Lipton 2004: 3) or "is probably true" (Josephson and Josephson 2003: 5; Psillos 2002: 614). The word "probably" in the quotes is used without any specific interpretation of probability theory, it merely acknowledges that the conclusion is not a deductive one (Psillos 2002: 614 fn. 17). Hence IBE is a strengthened form of abduction. IBE adds one premise to the abductive form of inference, the addition of which permits us, supposedly, to infer not merely a possible, but a true conclusion.

Peirce stated that abduction begins from a surprise, i.e., then "the surprising fact, C, is observed" (Peirce 1934: 5.189). Thus, neither abduction nor IBE can be conducted if there is no surprising fact to be explained. Aliseda (1997: 28) proposed the term "an abductive trigger" to stand for the surprising phenomenon that triggers abduction and needs an explanation. According to her, the concept of the abductive trigger is a relative one. What is surprising for one person might not be surprising for somebody else. These persons can have different background knowledge. Therefore, whether something is an abductive trigger depends on the particular content of the relevant background knowledge (BK). A fact E can be an abductive trigger only if the relevant background knowledge does not explain E, i.e., only if $BK \not\Rightarrow E$, where $\Rightarrow$ stands for an appropriate logical or explanatory relation. Depending on whether an abductive trigger is consistent with the background knowledge or not, Aliseda distinguishes between novel and anomalous abductive triggers (Aliseda 1997: 28; 2006: 47). A novel abductive trigger is one that is not explained by background knowledge, $BK \not\Rightarrow E$, but is compatible with it $BK \not\Rightarrow \neg E$. An abductive anomaly occurs when the abductive trigger contradicts the background knowledge,

i.e., when background knowledge explains the negation of the abductive trigger, $BK \Rightarrow \neg E$. Nevertheless, the existence of an abductive trigger does not compel one to seek its explanation. The idea is rather that one is in need of abduction only if, firstly, she stumbled upon an abductive trigger and, secondly, she wants to understand it.

Background knowledge is thus essential for assessing whether something is or is not an abductive trigger. If somebody does not have any background knowledge, then for this person everything she perceives is an abductive trigger. For an omniscient creature, on the other hand, abductive triggers cannot exist. According to Gabbay and Woods (2005: 87), "something is an abductive trigger when a memory search induces the agent to realize (usually tacitly) that the attendant cognitive irritation cannot be removed by what he currently knows." Gabbay and Woods put forward two conditions similar to Aliseda's novelty and anomaly that characterize an abductive trigger. It has to be unexpected and uncharacteristic. The unexpectedness of an abductive trigger means that it is something that "could not have been forecast solely on the basis of what one knew at the time" (Gabbay and Woods 2005: 197), i.e., something is an abductive trigger only if it is not explained (was not predicted) by background knowledge. The uncharacteristicness means that "in spite of the fact that its occurrence is now known, it presents the agent with an additional cognitive target which cannot be hit with what is now known" (Gabbay and Woods 2005: 198). In other words, an abductive trigger is uncharacteristic because, even though it is now a part of background knowledge, this background knowledge lacks the knowledge that would make the abductive trigger expected.

What follows from background knowledge cannot be an abductive trigger. If background knowledge contains a satisfactory explanation for some particular phenomena, if one knows (or at least thinks one knows) the causes of the phenomena, then there are no explanation seeking why-questions that could be asked about the phenomena. There is no rationale in explaining something that is already known. According to Thagard, it is "pointless to waste mental resources on something ordinary or expected" (Thagard 2007a: 227) or, according to Aliseda, "non-surprising facts should not be candidates for abductive explanations" (Aliseda 2006: 47). A known fact can become an abductive trigger only if one realizes that background knowledge possesses no explanation for it or that a known explanation is not actually satisfactory. The concept of anomaly in Kuhn (1996) stands for a genuine kind of abductive triggers.

Something is an abductive trigger only if it was not expected before it has been observed. However, an abductive trigger becomes a part of background knowledge after it has been observed. An abductive trigger in any instance of abduction or IBE has to be handled as true. It cannot be false. If it were false, then there would be no need for an explanation of it. Therefore, it always has to be considered as having a truth value "true" and probability $Pr(E) = 1$. In this respect, the abductive trigger acts similarly to old evidence and may bring the same problems as the problem of old evidence.

The appearance of an abductive trigger raises an abduction problem. The task here is to find an explanation that would modify background knowledge so that the background knowledge would accommodate the abductive trigger. As Lipton (2004: 21) puts it, the question is "what has to be added to knowledge to yield understanding?" A solution to an abduction problem is either to add

new propositions to background knowledge, to subtract propositions from it, or to modify background knowledge by first subtracting some propositions and then adding some other propositions. The only requirement to solve an abduction problem is to find or propose a link between the abductive trigger and background knowledge. There can be multiple solutions to an abduction problem and there is no talk yet about the goodness of an explanation.

An abduction problem is solved when a surprising fact ceases to be an abductive trigger and becomes derivable from background knowledge. According to Psillos,

> whatever the formal details of an act of explaining, it should incorporate the explanandum into the rest of the reasoner's background knowledge by providing some link (even by breaking a link) between the explanandum and other hypotheses that are part of the reasoner's background knowledge. (Psillos 2007: 445)

Or as Day and Kincaid put it,

> what makes an event expected is that it fits well with what else one knows. This tie between IBE, unification and overall coherence of belief is so frequent and pervasive that 'explains' and 'fits with' are often used synonymously. (Day and Kincaid 1994: 276)

The term "to derive" is often used instead of "to explain" in the context of abduction. It can stand for logical entailment, material implication or probabilistic dependence. Conclusion of an abductive inference is one which is possible, probable or worthy of conjecture. Abduction does not require an inference to the sole conclusion. Therefore, an abduction problem would be solved by any relation that makes an abductive trigger plausible. This relation

can be satisfied by any consequence relation. An abductive trigger in this relation is the consequent. The truth value of any abductive trigger is *true* (or probability $Pr(E) = 1$). Hence, abduction is satisfied by any consequence relation, because an abductive trigger as a consequent cannot obtain a smaller truth value than any of the possible antecedents. According to Aliseda,

> abduction is not one specific non-standard logical inference mechanism, but rather a way of using any one of these. (Aliseda 2006: 47)

Even though Wesley Salmon declares that without an exact explication of the concept of explanation IBE "has no clear meaning" (Salmon 2001a: 68), the theories of IBE are neutral about the question of what exactly the relation "to explain" stands for. According to Lipton,

> whether or not explanatory considerations are a guide to inference does not depend on whether we have an adequate account of explanation, any more than our use of a grammar to understand our language depends on our ability to give an adequate explicit account of the structure of that grammar. Moreover, if in fact we do use explanatory considerations as a guide to inference, to say that we do so seems to me not a meaningless or even a trivial claim, even in absence of an account of explanation, because we have some semantic grip on the concept of explanation in the absence of such an account. (Lipton 2001a: 100)

According to Psillos,

> I think that the very possibility of Inference to the Best Explanation as a warranted ampliative method must be

> examined independently of specific models of the explanatory relationship between hypotheses and evidence. Ideally, IBE should be able to accommodate different conceptions of what explanation is. This last thought implies that abduction (that is, IBE) is not usefully seen as a species of ampliative reasoning, but rather as a genus whose several species are distinguished by plugging assorted conceptions of explanation in the reasoning schema that constitutes the genus. (Psillos 2002: 606)

Similar declarations can also be found in Thagard (2007b) and (Newman 2009). Later in this work it will be claimed that different explanatory virtues can stand for different concepts of explanation or the depth of an explanation can be a place-holder for different concepts of explanation.

Notwithstanding the neutrality about the concept of explanation, the theories of IBE have a slight preference for the causal-mechanical account of explanation. For example, (Lipton 2004: ch. 3) devotes a whole chapter of his book to enumerate and defend the advantages of the causal account of explanation. Thagard also states that "in accord with much recent work in the philosophy of science, I hold that to explain a phenomenon is to describe a mechanism that produces it" (Thagard 2007b: 38).

Only after solutions to the abduction problem are proposed there arises a need to identify the best of these solutions. As Lipton puts it,

> there is always more than one possible explanation for any phenomenon [...] so we cannot infer something simply because it is a possible explanation. It must somehow be the best of competing explanations. (Lipton 2004: 56)

This best explanation, according to the theories of IBE, should be treated as the actual story about the abductive trigger.

There is only one additional requirement that the best explanation has to satisfy in order to be proclaimed true. A conclusion of IBE should be accepted not only if it is the best explanation, but, in the first place, if it is a good enough explanation. A question arises how to define the notion of a "good enough explanation." A practical and convenient answer to this problem is to treat any conclusion of abductive inference as a good enough explanation. The term "good enough" stands for "satisfactory for its purpose." The task of IBE is to find the best hypothesis out of hypotheses that do indeed explain the abductive trigger. Hence, the mere ability to explain the abductive trigger should be enough for a hypothesis to be considered as a good enough explanation and this, consequently, means that any abductive conclusion should be treated as a good enough explanation. If something does not explain the abductive trigger, it should not be accepted as an abductive conclusion at all. This account of "good enough" does not disturb the description of IBE as a two-step process: abduction plus evaluation of explanatory power. On the other hand, if, being a mere abductive conclusion, a mere ability to explain the abductive trigger is not good enough, then nothing is. Any criteria of "good enough" in addition to the mere abductive capacity would require evaluation explanatory goodness, i.e., of explanatory power, and (1) this evaluation would require at least two hypotheses (a sole hypothesis would not have a benchmark to evaluate it against), (2) this evaluation would already constitute IBE and (3) it would be an arbitrary point in a continuum of possible values of explanatory power.

The relation between abduction and IBE is generalized by the difference between a potential and the actual explanation. A conclusion

of abduction is a potential explanation: a possible, but unknown-whether-true, explanation. A conclusion of IBE is, allegedly, the actual explanation: the one of the potential explanations (one of the abductive conclusions) that is the true explanation. Because of that, every conclusion of IBE is a proper conclusion of abduction, but not every conclusion of abduction is a proper conclusion of IBE. In other words, every conclusion of IBE is an explanation, but not every explanation is the best explanation.

Abduction provides a pool of potential explanations. This is the first step of IBE. The next obvious question is how to determine which one of the potential explanations is the best one. The next two sections deal this this question.

1.3 Constitution of Explanatory Power: Explanatory Virtues

1.3.1 Explanatory Virtues

If explanatory power is a reason to accept a hypothesis, then the greatest explanatory power is an even better reason for that (Gabbay and Woods 2005: 101). Therefore, the next task in explicating IBE is to describe what the explanatory power as "explanatory considerations" or "explanatory loveliness" stands for. What are the criteria that discriminate that one hypothesis is a better explanation than another?

Explanatory virtues stand for these criteria. Explanatory virtues are sometimes also referred as "cognitive", "epistemic", "inferential" or "theoretical", presuming that these terms are synonymous. In this thesis the term "explanatory virtues" denotes the virtues that are used to judge the explanatory power. The terms "epistemic" and

"theoretical" are also used to denote particular kinds of explanatory virtues.

Explanatory virtues are the criteria according to which one compares the explanatory power of competing hypotheses. A hypothesis with the highest degree of explanatory virtues is declared the best explanation. The explanatory virtues act in the contexts of both discovery and justification, by guiding the choice of an explanation and by boosting or lowering the degree of confidence one can have in a particular explanatory hypothesis.

The explanatory virtues most often mentioned are explanatory coherence or, simply, coherence (with background knowledge), simplicity and unification. According to Lipton (2004: 67, 149), explanatory virtues play a role in, firstly, generating potential explanations and, secondly, in the choice of the best one out of them. Thus, the best explanation is a hypothesis that exhibits the ultimate degree explanatory virtues:

> best is not directly a judgment of truth but instead a summary judgement of accessible explanatory virtues. (Josephson and Josephson 2003: 15)

or

> explanatory coherence is a vehicle through which an inference is performed and justified. (Psillos 2002: 619)

One can notice a sign of epistemic rationalism again in a claim that explanatory considerations alone, without any appeal to confirmation or empirical testing, are sufficient for accepting a hypothesis:

> if a hypothesis has been chosen as the best explanation, then it has fared best in an explanatory-quality test with its competing rivals. So unless there is reason to think that

> it is superseded by an even better explanation, or unless there is reason to believe that the recalcitrant evidence points to one of the rivals as a better explanation, to stick with the best explanatory hypothesis is entirely reasonable. (Psillos 2002: 622)

Because IBE is legitimate even without any confirmatory considerations, it can facilitate a choice of empirically equivalent hypotheses (e.g., Day and Kincaid 1994: 275; Bird 2005: 8; McMullin 1996: 21; Psillos 1999: 170–174). McMullin even suggests labelling some of the non-empirical explanatory virtues as "complementary", because they can help to make a theory choice when empirical virtues are not able to do that.

Following Barnes (1995: 273 fn. 4) one could argue that aesthetic virtues (beauty, elegance) should not be conflated with explanatory virtues in the context of IBE. Even though the term "explanatory loveliness" has rather aesthetic connotations, its denotation in the context of IBE is the provision of understanding. Explanatory virtues as the sign of explanatory loveliness should provide understanding rather than aesthetic pleasure. On the other hand, when it comes to explicating the meaning of beauty or elegance in the theory choice, simplicity, for example, is both one of the most often cited explanatory virtues and one of the most often cited features of beauty. Hence, aesthetic virtues should be claimed to be derivative from some of the explanatory virtues. Moreover, it is difficult to explicate aesthetic virtues, because, on the one hand, they are ascribed features such as simplicity, symmetry, regularity, visualizability and, on the other hand, they are ascribed contrary features—complexity, diversity, asymmetry, contingency, abstractness—at the same time (Kuipers 2002: 299). Due to these reasons (they are derivative and too vague) we will exclude aesthetic virtues (beauty and elegance)

from the set of explanatory ones. However, the way they are analyzed and conclusions drawn about them are applicable to and will be used further in the analysis of other explanatory virtues.

An identification that a hypothesis predicts the abductive trigger or that a hypothesis is confirmed by the relevant empirical data is an empirical explanatory virtue. Theories possessing empirical virtues constitute the aim of the explanatory activity, i.e., an accumulation of theories that successfully accommodates the intended data. All other explanatory virtues are non-empirical or theoretical. The idea behind IBE is that explanations possessing the ultimate degree of theoretical virtues would have to exhibit the ultimate degree of empirical virtues, i.e., they would have to be confirmation-conducive (this does not preclude empirical virtues contributing to the evaluation of the explanatory loveliness).

An explanatory virtue that is truth-conducive in virtue of its logical form is a logical virtue. Logical explanatory virtues have to be truth-conducive in every possible world. If all the explanatory virtues were logical, then IBE would be a deductive form of reasoning. However, this is not the case. All other non-logical explanatory virtues are metaphysical, because if they are truth-conducive that can be only because the ontological structure of the world is such that allows these virtues to be truth-conducive. As Lipton states,

> a pattern of non-demonstrative inference that generally takes us from truth to truth in this world would not do so in some other possible worlds. (Lipton 1993: 101)

IBE is one of those patterns of non-demonstrative inference, therefore at least part of the virtues it relies on are of metaphysical character.

Epistemic virtues stand for explanatory virtues that are truth-

conducive. Explanatory virtues are epistemic virtues if they are truth-conducive. All the empirical or logical virtues are epistemic because they constitute the goal of the epistemic enquiry and explicate the concept of truthlikeness. For example,

> the requirements of internal consistency or predictive accuracy are prized not because they have previously been witnessed to accompany verisimilitude but because they are the elements of an explication of that very concept [...]. It remains of course possible for indicators of truth to be inductively learned by a scientific community but this is irrelevant to the a priori logical status of such criteria. (McAllister 1989: 38)

Pragmatic virtues are those that are useful or convenient for a human action. For example, van Fraassen (1980) claims that only empirical virtues are epistemic and all the theoretical virtues are at most pragmatic and that we should abstain from evaluating their epistemic status. For pragmatists, all and only the pragmatic virtues are epistemic. The theories of IBE, on the other hand, do not really concern themselves with whether explanatory virtues are pragmatic. The theories of IBE are interested only if explanatory virtues are epistemic, because IBE is truth-conducive if and only if the explanatory virtues, especially theoretical and metaphysical, are epistemic virtues. Whether explanatory virtues are pragmatic or not is irrelevant for IBE, the positive answer to this question at most can mean a useful or pleasant bonus for human practical life.

How do theoretical and metaphysical explanatory virtues, which do not constitute the goal of epistemic enquiry or define the concept of truthlikeness, come to be known and claimed to be epistemic? Researchers of the aesthetic virtues in the theory choice (McAllister

1989; Kuipers 2002) suggest that this is done by meta-induction, i.e., inductively identifying common non-empirical features in successful theories. For example,

> one could cast an inductive eye over the history of science and determine whether theories demonstrating certain forms of simplicity had as a matter of contingent fact tended to be closer to the truth—as this was later revealed—than other theories. (McAllister 1991: 10)

This does not imply that empirical or logical virtues cannot be discovered inductively, these virtues just have to have their separate justification (this can also be seen in the above quote from (McAllister 1989: 38)). However, the epistemic character of genuine theoretical and metaphysical virtues cannot be justifiable by any other means than empirical.

Different explanatory virtues are analyzed in the literature on IBE, abduction, explanation and theory choice. They are distinguishable into groups that stand for the breadth, depth, coherence, simplicity and empirical adequacy of an explanation.

Coherence. Coherence (also congruence, consonance, fit with background knowledge and plausibility in relation to background knowledge) is often claimed to be the most important of the explanatory virtues. It stands for the coherence between the explanatory hypothesis and relevant background knowledge. A hypothesis is better the more coherent it is. Background knowledge is a body of knowledge that is taken for granted when discussing a particular problem. This body of knowledge by itself might not be uncontroversial, but when used as the background knowledge it has to be assumed to be true or at least very close to truth:

> if some parts of the background knowledge are called into

> question, they do no longer belong to the background knowledge. (Kamps 2005: 318)

Background knowledge does not have to include all known facts, but it should include the abductive trigger. The abductive trigger has occurred, it is known fact, its is true; hence it should be treated the same way as background knowledge. If an abductive trigger is taken to be the part of the background knowledge then coherence with background knowledge would also mean coherence between an explanatory hypothesis and the abductive trigger.

One the one hand, coherence with background knowledge is sometimes referred to as an ordinary explanatory virtue. One the other hand, it is also referred to as (1) the most important explanatory virtue sometimes even sufficient to single out the best explanation (Harman 1968: 531; Psillos 2002: 615), (2) the starting point of any explanatory evaluation (picks out the set of potential explanations) (Lipton 2004: 151; Psillos 1999: 219) or (3) the evaluator of the relevance of other explanatory virtues (Lipton 2004: 139–140; Psillos 2007: 443). Even van Fraassen, who denies the epistemic claims about IBE, maintains in his pragmatic theory of explanation that an evaluation of explanatory answers as good or better

> proceeds with reference to the part of science accepted as 'background theory' in that context. (van Fraassen 1980: 141)

Other explanatory virtues might not even be required for determining the best explanation. Psillos (1999: 219), for example, thinks that other explanatory virtues are called for to assist only when the background knowledge cannot determine the best explanation on its own. Glass (2007) proposes to measure explanatory loveliness solely by a probabilistic measure of coherence.

Coherence with background knowledge is the most important of the explanatory virtues in the sense that it is the only explanatory virtue that is necessary and sometimes even sufficient for determining the best explanation. The background knowledge contains all the premises of every instance of IBE (the abductive trigger and information on why a particular hypothesis is the best one). If there is no background knowledge, then there are no premises to begin an inference with, thus there is no IBE. In this sense the background knowledge is an immovable obstacle for the proof of the formal validity of IBE. Formal validity of IBE would make conclusions of IBE true in every possible world. However,

> considering all possible models means testing in a knowledge vacuum. (Douven and Horsten 1998: 316)

The knowledge vacuum forbids any application of IBE and by the same token precludes the possibility of proving the formal validity of IBE.

The significance of the background knowledge is only questioned in cases of so-called creative abduction. Schurz (2008: 218) describes hypothetical (common) cause abduction as the most fundamental kind of creative abduction which does not assume any background knowledge except the knowledge about the abductive triggers. Creative abduction, which introduces new concepts, models, explanations, etc. is usually distinguished (Magnani 2001; Schurz 2008) to contrast it to the selective abduction, when one seeks an explanation in the pool of known explanations, i.e., in the pool of the background knowledge. But this does not mean that the background knowledge is not necessary in the cases of creative abduction. In these cases the background knowledge is simply insufficient to provide a good enough explanation. According to Gabbay and Woods,

> in solving abduction problems, the demands of originary (or creative) thinking are proportional to the depth of the abducer's ignorance, (Gabbay and Woods 2005: 64)

i.e., the demand for creativity is proportional to the lack of background knowledge. In an instance of novel abduction one cannot account for the abductive trigger with the background knowledge at hand, but the background knowledge is still operative, because the novel explanation should not contradict background knowledge and should be as coherent as possible with it.

Even if coherence with background knowledge is the most important explanatory virtue and background knowledge is taken for granted as true this does not mean that explanatory hypotheses cannot contradict and eventually alter background knowledge. Examples of the most radical alternations to the background knowledge are paradigm changes (Kuhn 1996) though less substantial adjustments can also be likely, for example, in cases of anomalous abductive triggers, i.e., when the background knowledge predicts facts that are contrary to abductive triggers. Nevertheless, in these kinds of situations an explanatory virtue of conservatism as a feature of coherence with background knowledge prefers the alternation to be as small as is sufficient to accommodate the abductive trigger or at least not bigger than in any of the alternative explanations. In this respect the explanatory hypothesis should correspond between relations in past and successor theory and explain past successes and failures of a predecessor theory.

Analogy is also a kind of coherence with background knowledge. Explanations that are analogous use background knowledge to form explanations that are similar to explanations already in use. Analogy resembles unification, but is not a kind of unification, because it does not use the same explanation, but only one that is similar. It is

a kind of coherence, because it only extrapolates background knowledge; thus making new knowledge not very different from knowledge already taken to be true. Thagard, for example, takes analogy to be one of the constitutive principles of explanatory coherence by stating that

> similar hypotheses that explain similar pieces or evidence cohere. (Thagard 2007b: 32)

Related to analogy is the virtue of understandability or intelligibility of an explanatory hypothesis when a new explanation uses familiar concepts and models or is illustrated by analogies with something known.

Consistency and intra-theory support stand for the internal coherence of an explanatory hypothesis. Causal, temporal, structural, etc. priority of explanans over explanandum (Huemer 2009a) should also be classified as a feature of internal coherence.

Theories of IBE do not explicate much what they mean by the concept of coherence. However, when using the very concept of coherence, proponents of IBE seem to go in the opposite direction to other studies in epistemology or philosophy of science. Probabilistic measures of coherence use probabilities of explanatory hypotheses as arguments to measure their coherence, but explanationists claim that coherence itself should be used to evaluate these probabilities. Conceptual explications of coherence use explanatory relations to evaluate coherence between a proposition and a belief-system, but theories of IBE use coherence between a proposition and a belief-system to evaluate explanations.

This strongly impairs the claims about IBE. Throughout the studies on coherence (Bartelborth 1999; BonJour 1985; Lehrer 1990; Lewis 1946; Rescher 1973; Thagard 1989) a rigid pattern emerges

that identifies two fundamental features of coherence for a set of propositions: (1) logical consistency or absence of contradiction, and (2) inferential-explanational relations. Mutual logical consistency is required as a necessary feature, but is an insufficient one. According to Bartelborth (1999), to believe that at one second "In front of me I see the White House", at the next "I see a BMW in front of the Empire State Building", and at third "I see the centre of Leipzig before me" is perfectly consistent, but the background knowledge does not allow us to claim that these beliefs are coherent. Thus something more than mere consistency is required to define coherence, because logically consistent propositions can become incoherent in relation to the background knowledge. Inferential-explanational relations in a set of propositions or, in the case of IBE, between the explanatory hypothesis and the background knowledge are claimed to constitute the second component of coherence. Lewis (1946: 338) talks about an inferential relation similar to conditional probability, which is not that strong as a deductive inference. Rescher (1973: 32–33) talks about connectedness. For BonJour (1985: 95–98) explanatory relations are one central element of coherence and their presence enhances the coherence of a system of beliefs. Bartelborth (1999: 220–221), Lehrer (1990: 95) and Thagard (1989: 436–437) define the second component of coherence more or less similarly: an explanatory hypothesis would cohere with background knowledge if it explains the background knowledge or if the background knowledge explains the explanatory hypothesis. These reciprocal explanatory connections stand for the slogan that coherence is a matter of "hanging together" (BonJour 1985: 93). This is consistent with the claim that the maximal possible coherence has to be the coherence of equivalent propositions (Bovens and Hartmann 2003; Fitelson 2003), because equivalent propositions entail one another.

If we stick to this definition of coherence then the central explanatory virtue of IBE becomes viciously circular or, as Day and Kincaid (1994: 277) put it, uninformative. The best explanation is the explanation most coherent with background knowledge. The explanation most coherent with background knowledge is the best explanation for the background knowledge or the proposition that is best explained by the background knowledge. Hence the claim that the best explanation is an explanation that is the most coherent with background knowledge becomes trivial and empty. The terms "the best explanation" and "the most coherent" denote the same thing. Further on in this text, a way out of this circularity will be proposed, according to which, other explanatory virtues, which are more precisely defined, explicate the concept of coherence.

Breadth. A hypothesis is better the more unifying it is. There are two main requirements for a hypothesis to be unifying, broad, consilient, have a wide scope, or explain a variety of facts. It has to account for at least two different kinds of facts (the more the better) and unification cannot be a simple conjunction of hypotheses or observational statements. Firstly, if a hypothesis successfully explains a very big number of different facts it makes all other competing hypotheses that explain only a few kinds of facts much less attractive. The more different kinds of facts a hypothesis explains or successfully predicts, the more unifying it is. Bartelborth (2002) calls this feature of unification the systematization force. Thagard provides the following two definitions of unification, where FT_i is the set of kinds of facts explained by T_i:

> (1) T_1 is more consilient than T_2 if and only if the cardinality of FT_1 is greater than the cardinality of FT_2;
> (2) T_1 is more consilient than T_2 if and only if FT_2 is a proper subset of FT_1. (Thagard 1978: 79)

According to Thagard these definitions are not equivalent, because the cardinality of T_1 can be larger than of T_2, even though T_2 can explain more important facts than T_1 and because of that be more preferable. This conflict can occur only when neither T_1 nor T_2 is complete—i.e., fail to explain all the facts of interest—although is not refuted by these facts. Thagard and Psillos (2002: 615) concur that when it happens the better explanation is one that explains more important facts. One can thus generalize that the value of unification brought by H_i is the cardinality of the set of relevant kinds of facts explained by H_i, i.e., the cardinality of the set of consequences of H_i. The larger the cardinality is, the more unifying the explanation given by H_i is. If the hypothesis successfully predicts new kinds of facts then unification increases. Therefore explanatory virtues referred to as the capacity to generate novel predictions, fecundity, fertility, fruitfulness and productive promise stand for the virtue of unification.

Secondly, an increase in the cardinality of the consequence set by a mere conjunction of different hypotheses or observational statements does not constitute genuine unification, because this does not add anything to what is known. Bartelborth (2002) calls this constraint on unification the organic uniformity and Schurz (2008) calls this the minimal adequacy criterion for second-order abductions, where the term "second-order abduction" denotes an explanation postulating the existence of a new kind of property or relation. H_i is unifying only if $k < n$, where n is the cardinality of the consequence set of H_i and k is the number of theoretical postulates of H_i. Unification requires a diversity at the end of the kinds of facts to be explained and a unity at the end of the explanation for these facts. According to Schurz,

> the introduction of one new entity or property merely for

> the purpose of explaining one phenomenon is always speculative and ad hoc. Only if the postulated entity or property explains many intercorrelated but analytically independent phenomena, and in this sense yields a causal or explanatory unification, it is a legitimate scientific abduction which is worthwhile to be put under further investigation. (Schurz 2008: 219)

This constraint on unification also accounts for why ad hoc hypotheses are not desirable in explanations. An ad hoc hypothesis is an opposite of unification, because it

> serves to explain no more phenomena than the narrow range it was introduced to explain. (Thagard 1978: 87)

An ad hoc hypothesis can be introduced into a theory only by conjunction with other theoretical postulates and it is the very thing the discussed constraint prohibits.

Myrvold (2003: 409–411) proposes a formal measure of the extent to which an explanatory hypothesis H unifies the set of evidence $\{E_1, \ldots, E_n\}$. It measures the extent to which H makes one piece of evidence yield information about other evidence. In other words, for Myrvold an explanatory hypothesis unifies a set of evidence if it makes different pieces of evidence informationally relevant to each other:

$$\begin{aligned} \text{Unification}_M(E_1, \ldots, E_n; H) &= I(E_1, \ldots, E_n|H) - I(E_1, \ldots, E_n) \\ &= log_2\left(\frac{Pr(E_1 \wedge \ldots \wedge e_n|H)}{Pr(E_1|H) \times \ldots \times Pr(E_n|H)}\right) \\ &\quad -log_2\left(\frac{Pr(E_1 \wedge \ldots \wedge e_n)}{Pr(E_1) \times \ldots \times Pr(e_n)}\right). \end{aligned}$$

McGrew (2003: 561–563) also suggests measuring the unification value or, to use his own term, the consilience of a hypothesis as an

extent of positive relevance between a set of evidence in the light of the hypothesis. According to him, a hypothesis H_1 is more unifying than H_2 if

$$\frac{Pr(E_1 \wedge \ldots \wedge E_n|H_1)}{Pr(E_1|H_1) \times \ldots \times Pr(E_n|H_1)} > \frac{Pr(E_1 \wedge \ldots \wedge E_n|H_2)}{Pr(E_1|H_2) \times \ldots \times Pr(E_n|H_2)}.$$

McGrew explicates the unification value of a hypothesis the same way as Myrvold does, i.e., as a difference in positive relevance that the unifying hypothesis brings. Schupbach (2005) shows that both of these account depend on the same inequality

$$Pr(E_1 \wedge \ldots \wedge E_n|H) > Pr(E_1|H) \times \ldots \times Pr(E_n|H)$$

and are thus identical in probabilistic terms. Hitchcock (2007: 438) also describes unification in terms of this inequality. Hence all three accounts imply that a value of the ratio

$$\text{Unification}_{MMH_i} = \frac{Pr(E_1 \wedge \ldots \wedge E_n|H_i)}{Pr(E_1|H_i) \times \ldots \times Pr(E_n|H_i)}$$

is sufficient to rank a set of j competing explanatory hypotheses, $\{H_i|1 \leq i \leq j\}$, according to the degree of unification they bring. The higher the value, the greater the unification is.

The MMH measure satisfies the claim about IBE that explanatory virtues can be brought to aid the choice between empirically equivalent hypotheses. For example, the two hypotheses H_1 and H_2 can have an equal likelihood $Pr(E_1|H_1) = Pr(E_1|H_2)$ on evidence E_1 alone and an equal likelihood $Pr(E_2|H_1) = Pr(E_2|H_2)$ on evidence E_2 alone. However, H_1 can make both pieces of evidence E_1 and E_1 more mutually positively relevant than H_2 and thus H_1 will have higher likelihood than H_2 on the pair of evidence E_1 and E_2, i.e., $Pr(E_1 \wedge E_2|H_1) > Pr(E_1 \wedge E_2|H_2)$. However, it does not satisfy the other desideratum of IBE that explanatory virtues should evaluate priors or likelihoods. The MMH measure uses likelihoods

in the determination of unification value, but according to the explanationists, these should be determined by the explanatory power and thus indirectly by unification too.

Lange (2004) claims that positive relevance is neither sufficient, not necessary to identify a genuine unification. Firstly, Lange (2004: 207–212) provides examplesm then Myrvold's measure identifies a unification even though there are no ontological-explanatory relations between the hypothesis and evidence. The measure captures too weak a sense of unification and is not sufficient, because it can attribute unification to hypotheses that create positive relevance between evidence, but are not cases of genuine unification. Unification is ascribed by the MMH measure independently of the properties and content of the hypotheses. Schupbach (2005) replies to Lange by claiming that a hypothesis can have the virtue of unification without being explanatory. Unification is only one of the explanatory virtues and sole unification does not constitute an explanation in advance. But Schupbach can reach such a conclusion only because he identifies unification with the positive relevance in advance and does not analyze whether the positive relevance is a good sign of unification. In one of Lange's examples the hypothesis H is "the two decay products annihilate as particle and anti-particle (so they must be electron and positron)", E_1 is "the decay product #1 is an electron" and E_2 is "the decay product #2 is a positron." H makes E_1 and E_2 mutually positively relevant. However, a fact stated in the hypothesis H is the result and not the cause of the corresponding evidence E_1 and E_2 and because of that H is not explanatory or ontologically prior to E_1 and E_2. It seems that by sticking to MMH as the measure of unification, one would have to accept any contingent correlating event as unification, if it brings sufficient positive relevance. This situation is similar to explaining a particular height

of a flagpole and the particular position of the sun by the particular length of the flagpole shadow (Salmon 1989: 47). The height of a flagpole and the position of sun are independent, but a particular length of the flagpole shadow will make the height of the flagpole and the position of the sun positively relevant. Maybe this dependence can be called unification, but we are considering unification as an explanatory virtue and the cases of non-explanatory unification are of no interest or relevance. Therefore, if a measure captures cases of unification that are irrelevant and does not capture the relevant cases then the measure is unsatisfactory for the present purpose: the explication of explanatory unification.

Secondly, Lange (2004: 212-214) provides an example of a unifying hypothesis (a common cause explanation) that does not bring any positive relevance in Myrvold's measure. The measure can be too demanding and the positive relevance is not necessary for unification, because a genuinely unifying hypothesis can derive a negative value of unification. According to Schupbach, the hypothesis in Lange's example cannot be a unifying explanation, because it is a common cause explanation and common cause explanations cannot be unifying in the sense accounted for by MMH. There are allegedly two different ideas of unification: one suggested by MMH, and unification as generality (e.g., a common cause explanation). But how can a hypothesis have a big enough consequence set and not be unifying? One answer is that there are different types of unification, but a more simple answer can be that MMH does not satisfy relevant intuitions about the concept. Some hypotheses in Lange's examples are claimed by Schupbach to be unifications, because the MMH measure identifies them that way. Some hypotheses are claimed are claimed to be 'not unifying', because the MMH measure does not identify them that way. However, the very problem under considera-

tion is whether the MMH measure is the measure of unification, but Schupbach argues for the MMH measure by taking it for granted as a genuine criterion of unification and claiming that if H does not comply with the MMH measure then it is not unifying. Schupbach's arguments seem to be rather ad hoc.

One of Lange's examples depicts a situation when the evidence set is positively relevant apart from the hypothesis H, but becomes statistically independent given H (and thus the MMH measure derives a negative value of unification). These conditions are similar enough for Reichenbach's Common Cause Principle to hold. Reichenbach's common cause principle (CCP) is also a model of unification (Reichenbach 1956: 157–167). According to CCP, when there is an unexpected correlation $Pr(A \wedge B) > Pr(A) \times Pr(B)$ between the events A and B, there exists a common cause C for these events that satisfies the following four conditions—the conjunctive fork—and thus screens off the correlation:

$$\begin{aligned} Pr(A \wedge B|C) &= Pr(A|C) \times Pr(B|C), \\ Pr(A \wedge B|\neg C) &= Pr(A|\neg C) \times Pr(B|\neg C), \\ Pr(A|C) &> Pr(A|\neg C), \\ Pr(B|C) &> Pr(B|\neg C). \end{aligned}$$

Even though for Schupbach common causes are not unifications, at least in the sense depicted by the MMH measure, for Schurz, for example, CCP is the leading principle of causal unification:

> an explanation of the three E_i by three distinct C_i is clearly inferior, because, in contrast to the common cause explanation, it cannot explain the correlations between the E_i—it rather shifts this problem into unexplained correlations between the C_i. (Schurz 2008: 222)

The MMH measure and CCP describe contradictory conditions for unification to occur. For the MMH measure the mark of unification is the transition from statistical independence to the positive relevance conditional on H. For CCP the mark of unification is the transition from positive dependence to statistical independence conditional on H. McGrew and Schupbach notice that every hypothesis satisfying CCP would be disunifying according to the MMH measure, but interpret this as an advantage rather than a disadvantage for the MMH measure.

According to the MMH measure, every hypothesis that entails its evidence cannot be unifying, because every H that entails its evidence makes its evidence mutually positively irrelevant conditional on H. This is a serious, if not a fatal, shortcoming of the MMH measure. If unification by the entailment is not unification, as proponents of the MMH measure claim (or at least it is another kind of unification), then what is the reason to suppose that unification by making positive relevance is the genuine unification? A possible answer may be that a hypothesis possessing this kind of unification gets better evidential support due to its unifying as making positive relevance power. But why, then, does Schupbach refuse to call unifying a hypothesis that has high evidential support and, if true, would explain several pieces of evidence, but does not bring positive relevance (as is depicted in one of Lange's examples)? It is probably because the MMH measure is neither necessary nor sufficient to do its task.

However, there is one more problem inherent to both the MMH measure and CCP. All hypotheses that entail their evidence have the same MMH value equal to 0, even if a hypothesis H_1 would unify only two different kinds of evidence and another hypothesis H_2 would unify ten different kinds of evidence. The MMH measure would

claim that H_1 and H_2 are equally unifying, but it is unintuitive, because H_2 unifies more kinds of evidence than H_1 does. This is applicable to any other set of hypotheses with the same MMH value, but different cardinalities of the sets of evidence they unify. This shortcoming is also applicable to a set of possible common causes satisfying the conjunctive fork, but unifying different numbers of correlating evidence.

Unification thus is not a matter of degree. A unifying hypothesis either unifies a particular number of distinct kinds of facts, or it does not. Therefore, the value of a unification measure has to be the number of explained kinds of facts. But what does decide whether a piece of evidence is explained by the hypothesis or not? The entailment of the evidence is a good candidate. Moreover, if a hypothesis provides only a statistical or probabilistic explanation, then the number of pieces of evidence made positively relevant or screened off should be counted as the value of unification. In other words, unification value of H should be $U = n$, where n is the cardinality of the set of evidence $\{E_1, \ldots, E_n\}$ that is entailed, made positively relevant or screened off by H.

However, the entailment, positive relevance and screening off do not always depict genuine explanatory relations. Moreover, there is still a feel of circularity, because the best explanation is, at least to some extent, the most unifying hypothesis, but the most unifying hypothesis is one that explains the most. These two problems can be escaped if the virtues that stand for the depth of an explanation provide criteria that determine if an explanation occurs.

Depth. A hypothesis is better the deeper it is. There are two main elements of the depth of an explanation. First, H_1 is better than H_2 if H_1 explicates a causal-nomological mechanism that produces the abductive trigger and H_2 does not. Secondly, H_1 is better

than H_2 if the mechanism posited by H_1 is deeper, more specific, precise, fundamental, informative or illuminating than one posited by H_2. How should one evaluate the depth of an explanation? According to Thagard,

> a deeper explanation for an explanatory mechanism M1 is a more fundamental mechanism M2 that explains how and why M1 works. M1 consists of parts, and M2 describes parts of those parts whose properties and relations change in ways that generate the changes in the properties and relations of the parts in M1. (Thagard 2007b: 38–39)

Hence, the degree m of the depth of a hypothesis can be measured as the cardinality of a set $\{M_i | 1 \leq i \leq m\}$ of explanatory mechanisms posited by the hypothesis, where M_i is a description of a more fundamental mechanism for or a structural elaboration of M_{i-1} if there is one.

For every explanation H_1, even the most plausible one, one can always ask for reasons H_2 why it should be the case that H_1 is true. Next, one can further ask what the reasons H_3 are why H_2 should be true and so on, ad infinitum. This is the so-called why-regress problem. Lipton (2004: 22) suggests that "explanations need not themselves be understood." This is definitely true. Somebody might have stolen my bicycle and this is a very good explanation why I cannot find it at the place where I left it. This explanation is fine even if I do not know why somebody wanted to steal it. However, if I knew why it was stolen, I would have additional reasons to believe in the hypothesis that it was stolen. Hence, even though a deeper explanation is not necessary to explain something, a deeper explanation—if there is one—by providing additional reasons, would be a better explanation than one that is not that deep.

A preference for the explanatory depth stems from the preference for causal-mechanical explanations that is exhibited by some proponents of IBE (e.g., Lipton 2004; Thagard 2007b). Because of that preference, the terms standing for the explanatory virtue of depth are most commonly associated with the causal mechanical account of explanation. Psillos, however, proposes not to stick to a particular concept of explanation. According to him,

> it seems to me methodologically useful to treat the reference to explanation in IBE as a 'placeholder' which can be spelled out in different ways in different contexts. [...] the general ways in which explanatory considerations can enter into defeasible reasoning can be specified without a prior commitment to the nature of the explanatory relation. (Psillos 2002: 606–607)

Hence, the depth of an explanation should be considered to be an ability of the explanation to be explained yet further, no matter what concept of explanation is employed. Moreover, different accounts of explanation can provide criteria to determine whether a genuine explanation occurs. An explanation should not be good enough if it does not provide any kind of explanation, i.e., if its depth is equal to 0.

Simplicity. A hypothesis is better the simpler or more parsimonious it is. Ironically, simplicity is the most complex of the explanatory virtues. Sober (2001), for example, talks about two kinds of simplicity (semantic and syntactic). Niiniluoto (1999a: 164) mentions four kinds of simplicity (ontological, syntactical, structural and methodological). Beebe (2009: 609) enumerates six different kinds of simplicity (one psychological kind, two ontological kinds, and three explanatory kinds). Simplicity is the most often-cited

aesthetic virtue of theories. It is also one of the most often-cited explanatory virtues, often explicated without any aesthetic connotations.

Since any hypothesis is expressed in a language any kind of simplicity has to be reducible to either semantic simplicity or syntactic simplicity. If someone says that a hypothesis is simple that means either the hypothesis expresses things that are simple or the hypothesis is expressed in a simple way. Semantic simplicity is basically ontological simplicity, known also as the principle of Ockham's razor: entities should not be multiplied beyond necessity. Ontological simplicity consists of minimizing the number of entities posited by a hypothesis: causes, laws, objects, principles, properties and other possible primitive explanatory notions. Hence, the term "entities" is taken "in the broadest sense" (Huemer 2009b: 216 fn. 1). Syntactic simplicity is structural simplicity. It consists of minimizing the number of structural components of the language of a hypothesis, or the number of structural components of a hypothesis itself: symbols, vocabulary, adjustable parameters, etc. Some structural components of a hypothesis—axioms, hypotheses, auxiliary hypotheses, theoretical postulates, etc.—can be considered to be the semantic and syntactic components of theories at the same time.

However, simplicity is not always desirable. Salmon (2001b: 129) states, for example, that in social sciences simple hypotheses are not always plausible, because they can be oversimplifications. Carruthers (2006: 151) claims that in the biological realm simplicity is implausible, because one should expect biological systems to be "messy and complicated, full of exaptations and smart kludges."

Simplicity is similar to unification in the sense that a hypothesis is better the less it takes to explain something. Unification asks to explain more facts with same resources. Simplicity asks to ex-

plain same facts with fewer resources. Therefore, informativeness (to provide more information with fewer resources) is a kind of simplicity too. Simplicity is desirable, because "by introducing sufficiently many 'hidden entities' one can 'explain' anything one wants" (Schurz 2008: 219). The preference for simplicity is also a means to avoid proliferation of trivial alternative theories constructed, for example, by disjunctive addition(s) of a contingent proposition. If H is the best explanation, simplicity prevents $H \vee A$ from being the best explanation as well. Simplicity prefers that no subpart of a hypothesis would make same explanations as the whole hypothesis does. Aliseda (2006: 71) notices that this leads to non-monotonicity. An addition of at least one proposition to the hypothesis can prohibit inferring a former inference of the hypothesis.

When it comes to the assessment of the degree of simplicity, syntactic simplicity is more difficult to evaluate. According to Sober (2001: 16), the same propositions can be expressed in many different ways, using different languages; therefore syntactical simplicity is not linguistically invariant. On the other hand, this problem is not a threat to the semantic kind of simplicity, because what a hypothesis states should not depend on the language it is stated in.

One may take simplicity to be a mere number of posited explanatory entities or the syntactic length of a hypothesis. However, this kind of measure would not be able to evaluate that a hypothesis that explains, for example, 10 kinds of facts with only 3 assumptions, is intuitively simpler than a hypothesis that takes 2 assumptions, but explains only 3 different kinds of facts. Simon (2001: 34) uses the term "simplicity" to stand for the length of the string of binary information. The simpler a hypothesis is the shorter the string. In other words, the shorter the syntactic length of the hypothesis is (simplicity is reciprocal of complexity). One the other hand, Si-

mon (2001: 35) uses the term "parsimony" to stand for the ratio of the complexity of the data to the syntactic length of the hypothesis and claims that parsimony thus defined is preferable to simplicity. According to him,

> parsimony brings simplicity in its wake; but simplicity in theory without parsimony in the relation between the theory and data is bought only at the price of weakening the goodness of approximation of our descriptions, narrowing the range of phenomena over which they extend, and impoverishing our understanding of the phenomena. (Simon 2001: 69)

Hence, Simon implies that the length of a string of binary data that explains many different kinds of facts (parsimony) will eventually be shorter (simplicity) than other strings that explain the same data, but not in a similarly unifying manner.

A similar measure of simplicity is proposed by Thagard (1993: 90). According to this measure, the simplicity of a hypothesis depends on the number of co-hypotheses or theoretical postulates it employs,

$$\text{Simplicity}_T = \frac{\text{facts explained by } H - \text{co-hypotheses of } H}{\text{facts explained by } H}.$$

The measure gets a value of 0 when it needs a co-hypothesis for every fact it explains and a value of 1 when in employs no co-hypotheses at all. There should be no worry about dividing by 0, because if a hypothesis does not explain anything, it will never be evaluated. The idea behind Thagard's measure is that simplicity has to minimize the number of additional assumptions needed to explain a particular amount of evidence. The worst hypotheses according to this measure are ones that need at least one additional assumption for

each different piece of evidence they explain. This idea is in a perfect match with Peirce's thesis that simplicity "adds least to what has been observed" (Peirce 1935: 6.479). In other words, Thagard's measure is higher the fewer additional assumptions are added in the hypothesis to explain evidence. However, Thagard's measure of simplicity discriminates hypotheses by the number of additional co-hypotheses they employ and cannot tell how simple the hypotheses are themselves without the co-hypotheses.

Simplicity is often discussed in relation to the curve-fitting problem. There are several information measures that incorporate syntactic simplicity (as the number of adjustable parameters or the syntactic length of a hypothesis) as one of the arguments: Akaike's Information Criterion, Bayesian Information Criterion, Minimum Description Length or Minimum Message Length. However, the arguments in relation to the curve-fitting problem are not relevant to the problem of IBE. In curve-fitting one seeks a curve that most adequately represents the relation between the dependent variable and independent variables. Hence, the causes (the independent variables) of the abductive triggers (the dependent variable) are assumed to be known and established and only the details of the dependence are of interest. The theories of IBE, on the other hand, claim that the simplest, but not yet known or established cause should be treated as the true one. Moreover, these information measures treat simplicity syntactically as the number of adjustable parameters or the amount of information that takes to represent a hypothesis. Hence they do not add anything new to the ongoing description of simplicity.

To sum up, there are two candidates for the measure of simplicity. The first identifies simplicity with k, where k is the syntactic length of a hypothesis or the number of explanatory entities. The second holds that simplicity is the ration $\frac{n}{k}$, where n is the consequence set

of the hypothesis or the number of explained different kinds of facts. The second one is intuitively more appealing.

Empirical Adequacy. Empirical adequacy constitutes the very aim of science (at least in the positivistic vein), thus it should be the most desirable feature in an explanation. However, empirical adequacy is not as much a feature of explanatory power, as it has to be the necessary effect of an inference that is the best explanation (e.g., Lipton 2001a: 109). In other words, if IBE is truth-conducive, empirical adequacy as an empirical virtue has to be the effect of other explanatory virtues.

Nevertheless, empirical adequacy and testing of predictions can enter explanatory considerations through coherence with background knowledge. Every piece of empirical data is simply a piece of background knowledge and every empirical refutation or inconsistency of some hypothesis can be considered as incoherence with background knowledge. This is how additional evidence can perform the eliminative function.

1.3.2 Relations Between Explanatory Virtues

Unification and Simplicity. The multiplicity of different explanatory virtues is thus reducible into groups of explanatory virtues that stand for coherence, breadth, depth, simplicity and empirical adequacy. The multiplicity can be reduced even further by describing the interrelations between the virtues. The most conspicuous connection among the explanatory virtues is between the unification and simplicity. Talk about simplicity can sometimes be substituted by the talk about unification and vice versa. These are often structurally and conceptually similar. Unification is maximality of the set of explained kinds of evidence and simplicity is minimality of the set of constituents of the hypothesis. Both can be measured in a similar

mathematical form. Schurz (2008: 219–229) states that an adequate and non-trivial causal or explanatory unification occurs only if postulated entity or property explains many phenomena. More precisely, then

$$n > k,$$

where n is the number of explained kinds of evidence and k is the number of theoretical postulates of H.

Simon (2001: 35) measures parsimony as the ratio of the complexity of the data to the complexity of a hypothesis. The hypothesis should be parsimonious if the complexity of it is smaller than the complexity of the data is explains. If k stands for the complexity of the hypothesis and n stands for the complexity of the data, then if the hypothesis is parsimonious the corresponding ratio should be higher than 1,

$$\frac{n}{k} > 1,$$

which is equal to $n > k$.

Thagard (1993: 90) measures simplicity as ratio

$$\text{Simplicity}_T = \frac{n-k}{n}$$

and the range of its value is $[0, 1]$. A hypothesis is simpler the higher above 0 its value. A hypothesis is simple, at least to some degree, if

$$\begin{aligned} \frac{n-k}{n} &> 0 \\ \frac{n(n-k)}{n} &> n \cdot 0 \\ n-k &> 0 \\ n &> k \end{aligned}$$

Hence, Schurz requires that a unifying hypothesis satisfies the same requirements as Simon requires of a parsimonious hypothesis or Thagard requires of a simple one. Unification favors hypotheses that

explain more facts with same resources. Simplicity favors hypotheses that explain same facts with fewer resources. Both unification and simplicity favor hypotheses that explain as many facts as possible with as few resources as possible.

Semantic simplicity is a straightforward form of unification. If H_1 and H_2 explain the same set of evidence, but H_1 poses fewer explanatory entities than H_2 then H_1 unifies the evidence with its lesser quantity of presuppositions. For example, according to Huemer,

> (H_1) is the simpler hypothesis, in so far as it postulates a single cause, while (H_2) postulates two independent causes. (Huemer 2009b: 225)

Hence, for Huemer a hypothesis postulating fewer causes (fewer explanatory entities) is the simpler one. A description of the same relation can also be found in Psillos:

> H^k, on the other hand, subsumes the explanation of all the data under a few hypotheses, and hence it unifies the explananda. (Psillos 2002: 616)

However, Psillos calls a hypothesis postulating fewer hypotheses (fewer explanatory entities) as unifying. Hence, philosophers use both terms "unification" and "simplicity" to refer to the very same phenomenon.

According to Forster and Sober (1994) unification and simplicity operate identically in the curve-fitting. Complex and disunified theories are more inclined to over-fit the data than simpler and more unified theories are. The more adjustable parameters a theory has or the more tailor-made to fit a particular piece of data the theory is, the lower the estimated predictive accuracy of the theory is.

In psychological experiments, in order to test whether people show preference for simpler hypotheses, more unifying hypotheses stand for simpler hypotheses. In an experiment by Read and Marcus-Newhall (1993) participants rated a conjunction of a set of narrow explanations, each narrow explanation individually and a unifying explanation that accounted for the same facts as the set of narrow explanations. Simplicity was claimed to be operative, because people rated the unifying explanation as better than a conjunction of narrow explanations. In an experiment by Lombrozo (2007), preference for simple explanations was also assessed by treating simpler explanation as one that provided a common cause for several pieces of evidence.

Unification and simplicity are often treated as similar. Earlier, two possible measures of simplicity were distinguished. The first identifies simplicity with k, where k is the syntactic length of a hypothesis or the number of explanatory entities. The second holds that simplicity is the ration $\frac{n}{k}$, where n is the consequence set of the hypothesis or the number of explained different kinds of facts. The second one is identical to unification. Therefore, following Ockham's advice not to multiply entities, unification will subsume the second measure and further simplicity will stand only for the first measure, i.e., the mere syntactic length or the number of posited explanatory entities.

Coherence and Other Explanatory Virtues. Conceptually, coherence between propositions is associated with explanatory relations between those propositions. But, according to the proponents of IBE, the power of an explanatory relation is to a major degree a function of coherence. Thus, the theories of IBE are stuck in a vicious circle. This circularity is a manifestation of a more general problem with coherence. As Olsson notes,

> the absence of a clear account has been noted as a troublesome fact ever since the days of the British idealists, and more recent coherence theories fare no better, in the lights of their critics, than their idealist ancestors did. [...] In the few cases where coherence theorists have actually proposed clear definitions, they can be seen, on closer scrutiny, to be incorrect. (Olsson 2005a: 13)

If coherence with background is to be the main explanatory virtue in IBE, it has to have a connection with other explanatory virtues. There has to be an explanation of why coherence is the main explanatory virtue or how other explanatory virtues contribute to coherence. Therefore, a way out of the circularity is an explication of coherence in terms of other explanatory virtues: unification, depth and simplicity. Coherence is the main explanatory virtue, because it generalizes the other virtues.

A hint in that direction can be found in Psillos (2002). According to him, explanatory virtues, which he calls "structural standards of explanatory merit",

> safeguard the explanatory coherence of our total belief corpus as well as the coherence between our beliefs corpus and a new potential explanation of the evidence. (Psillos 2002: 616)

However, Psillos is not explicit in the details as to how exactly particular explanatory virtues contribute to the coherence. His only idea about the relation between coherence and other explanatory virtues is that explanatory virtues operate when background knowledge cannot sufficiently discriminate the best explanation.

One of the most straightforward explications of coherence by means of other explanatory virtues can be found in Thagard (2007b).

He claims that explanatory coherence of a hypothesis can be increased over time in two main ways. Either, when the hypothesis explains new facts, i.e., by providing unification (broadening), or when the hypothesis and its success are explained by yet another hypothesis, which has its own independent evidential support, i.e., by providing a deeper explanation. Hence Thagard (2007b) explicates coherence of a hypothesis as the function of the breadth and depth on the hypothesis.

Coherence is sometimes claimed to be a wide concept that includes the narrower concept of unification. For Schurz (1999), firstly, the total coherence of a set is a function of the coherence of its parts. Secondly, coherence is constituted by inferential relations between propositions. Thirdly, coherence is circular: every proposition in a coherent set is inferable from the rest of the propositions in the set. Unification, on the other hand, for Schurz means inference of as many phenomena as possible from as few basic phenomena as possible. Hence, unification here is not symmetric. From all this it follows that

> coherence minus circularity equals unification. (Schurz 1999: 98)

Unification for Schurz is thus a constituting element of coherence. Bartelborth (1999), too, argues for the very same relation between coherence and unification. Firstly, coherence between phenomena for Bartelborth is constituted out of the explanatory relations between the phenomena. Secondly, to explain the phenomena is to establish connections between them. Thirdly, unification operates by establishing explanatory connections between different phenomena. Hence, unification creates explanatory connections between the phenomena and thus creates coherence between them. In his own

words,

> it should at least be obvious by now that the unification approach harmonizes very well with the coherence account of justification. Good explanations in the unification sense create substantial connections between our observational beliefs. (Bartelborth 1999: 218)

The same intuition relating coherence and unification can be derived from the definition of coherence by Shogenji (1999) and the MMH measure of unification taken together. According to Shogenji, the degree of coherence of a set of evidence $\{E_1, \ldots, E_n\}$ is equal to

$$C_S(E_1, \ldots, E_n) = \frac{Pr(E_1 \wedge \ldots \wedge E_2)}{Pr(E_1) \times \ldots \times Pr(E_2)}.$$

According to the MMH measure of unification, the degree of unification H_i brings is equal to

$$\text{Unification}_{MMH_i} = \frac{Pr(E_1 \wedge \ldots \wedge E_n|H_i)}{Pr(E_1|H_i) \times \ldots \times Pr(E_n|H_i)}.$$

If both measures were true (however, this seems unlikely, because both have unintuitive consequences), it would mean that the coherence of a set of evidence is the positive relevance between that evidence and the unification is the positive relevance given a hypothesis. Hence, something would become coherent if it could be unified. Unification would invoke coherence.

The degree of simplicity is also sometimes positively associated with the degree of coherence. Thagard is one of boldest examples in that respect. Even though he does not deny that complex hypotheses can sometimes be preferable, he maintains that

> the more hypotheses it takes to explain something, the lower the degree of coherence. (Thagard 2007b: 32)

Bovens and Hartmann (2003), on the other hand, provide an example that contradicts the Thagard's claim. Adding propositions to the set may increase the coherence of the set:

> certainly the information pair $S = \{$[My pet Tweety is a bird], [My pet Tweety cannot fly]$\}$ is less coherent than the information triple $S' = \{$[My pet Tweety is a bird], [My pet Tweety cannot fly], [My pet Tweety is a penguin]$\}$. (Bovens and Hartmann 2003: 29)

Simon (2001: 69) also maintains that syntactically simpler hypotheses tend to be less coherent, because the shorter the string of data describing something the less accurate the description can be, the narrower the range of phenomena it explains and the less understanding it provides. Hence, the longer the syntactical structure of a hypothesis, the more detailed and accurate information one expects to get from it. A longer string of information has more ways to be coherent with background knowledge than a shorter one. Moreover, the same holds for semantic simplicity. If a hypothesis is coherent with background knowledge—i.e., it has many explanatory relations with it—then it should employ the same entities that reside in the background knowledge. The more entities from the background knowledge the hypothesis employs, i.e., the more the hypothesis is semantically complex, the higher coherence one expects.

On the other hand, adding too much complexity with the aim of increasing coherence can be a mere ad hoc endeavor to salvage a hypothesis. A too complex hypothesis can become very coherent only with a very small part of background knowledge it intends to explain, and loose links with the rest of the knowledge pool. Hence, some complexity can contribute to coherence at first, but it can subtract coherence if the complexity gets too big.

Coherence is claimed to consist of consistency and mutual explanatory relations. Coherence of an explanatory hypothesis with background knowledge thus consists of its consistency with background knowledge and mutual explanatory relations between the hypothesis and background knowledge. Provided that the hypothesis provides a genuine explanation of its abductive trigger, the explanatory relations are stronger the deeper the explanation provided is and the more different kinds of background knowledge facts it explains together with the abductive trigger (the hypothesis explains background knowledge together with the abductive trigger), and vice versa. The explanatory relations are stronger the more detailed information can be provided by background knowledge as to why the hypothesis can be true and if the hypothesis uses patters of explanations already found in background knowledge (background knowledge explains the hypothesis).

The hypothesis explains background knowledge	**Background knowledge explains the hypothesis**
The deeper the explanation provided by the hypothesis is	The more detailed information can be found in background knowledge in support of the hypothesis
The more different kinds of background knowledge facts the hypothesis explains together with the abductive trigger	The hypothesis uses patterns of explanations already found in background knowledge

The kind of simplicity that is structurally similar to unification (that is here subsumed under unification) is a constituting part of co-

herence. However, the kind of simplicity— like the syntactic length or the number of posited explanatory entities—can be inversely proportional to the degree of coherence. Thus the latter kind of simplicity does not enter into the above mentioned description of coherence.

Even if unification, depth or simplicity did not constitute parts of coherence, unification and simplicity are at least considered by some philosophers to be the constituting parts of explanatory power. Thagard (1993: 91) computes the explanatory value of H as

$$\text{Explanatory power}_T(H) = \text{Simplicity}(H) \times \text{Unification}(H).$$

Carrier explicates explanatory power too as a function of unification and simplicity:

> Theories with great explanatory power need a minimum of independent principles to account for a broad class of phenomena in an accurate fashion. (Carrier 2009: 198)

However, by 'simplicity' both Thagard and Carrier mean the same thing we subsumed under unification. Therefore, for them the explanatory power is really the function of unification. Moreover, even though the depth of an explanation is not mentioned here it is implicit in the judgement of the adequacy of a hypothesis. No unifying or simple hypothesis would be explanatory if it is not deep enough, i.e., if it does not provide any kind of explanation. Therefore the depth of an explanatory hypothesis, for example, in Carrier's quote above should stand for "to account in an accurate fashion."

To sum up, the explanatory power of a hypothesis is measured by the degree of explanatory virtues it exhibits. The best explanation is the most explanatorily virtuous one. This is the second step of IBE. Coherence, unification, explanatory depth and simplicity are considered to be the main explanatory virtues. However, if we want

to define coherence in a non-question-begging way, we have to define coherence not as a primitive explanatory virtue, but as derivative from the rest of explanatory virtues. Coherence is constituted from two elements: consistency and explanatory relations. Thus defined, the concept of coherence appears to be synonymous with the concept of explanatory power: the best explanation is the explanation that is the most coherent with background knowledge, i.e., consistent, most unifying, the deepest, and the simplest.

1.4 Irrelevance of Probabilistic Considerations

1.4.1 Probabilistic Accounts of Explanatory Power

Explication of the explanatory power as a function of explanatory virtues (the explanationist account) is not the only one in philosophical literature. There are also probabilistic accounts of explanatory power. Probabilistic modeling of scientific or ordinary reasoning bears the name of Bayesianism. This section is going to analyze the probabilistic or Bayesian accounts of explanatory power and claim that the explanatory virtues account of explanatory power, by being more fundamental, is superior to the probabilistic ones.

In probabilistic terms the best explanation is often associated with a hypothesis that provides the highest likelihood for an abductive trigger, $Pr(E|H)$, i.e., the maximum probability of an abductive trigger given an explanation is true (maximum likelihood account, ML). For example, according to Okasha, distinction between explanatory loveliness and explanatory likeliness (Lipton 2004) can correspond, respectively, to the terms $Pr(E|H)$ and $Pr(H)$. More particularly,

> to decide whether a given explanation of a phenomenon is

> lovely, we ask the question: if it were true, would it render the phenomenon intelligible? (Okasha 2000: 704)

Lipton (2001a: 110) himself agrees that this identification is a tempting one, but denies that it is correct. The main reason for the skepticism is that a hypothesis may provide an abductive trigger with a high probability or even entail it without explaining it. Counterexamples to the deductive-nomological model of explanations illustrate this: for example, a particular length of the flagpole shadow and a particular position of the sun would entail the particular height of the flagpole, but, intuitively, would not explain it (Salmon 1989: 47). Huemer (2009a) claims that this problem can be escaped if the explanans is explanatorily prior to the explanandum. Given this condition,

> Bayes' Theorem seems to provide at least partial support for the explanationist approach: in choosing between candidate explanations h_1 and h_2 for evidence e, one factor that seems relevant is the likelihood ratio $P(e|h_1)/P(e|h_2)$. The greater this is, the better h_1 is as an explanation of e, compared to h_2: other things being equal, the hypothesis that more strongly predicts the evidence is the better explanation. (Huemer 2009a: 353)

Niiniluoto (2004: 72–73) also highlights ML as a plausible probabilistic account of IBE.

McGrew (2003) proposes a measure of explanatory power that is mainly a function of likelihood. McGrew derives his measure from Peirce's definition of abduction, in which an abductive trigger stands for "a surprise" and "to explain" stands for "to be a matter of course" (Peirce 1934: 5.189). For McGrew the explanatory power is the extent to which an explanatory hypothesis transforms

a surprising fact into a matter of course:

$$E_M = \frac{Pr(E|H)}{Pr(E)}.$$

According to this measure the value of explanatory power does not change even if new irrelevant evidence is added to the evaluation. The value does not decrease and thus the measure provides a very counter-intuitive result (Schupbach and Sprenger 2011). Moreover, McGrew's measure cannot assess an explanation of anomalous abductive trigger. There is a danger of division by zero. If an abductive trigger is anomalous, then $BK \vdash \neg E$, this implies that $Pr(\neg E) = 1$ and, consequently, that $Pr(E) = 0$.

Schupbach and Sprenger (2011) provide their own probabilistic measure of explanatory power. It is an elaboration of McGrew's measure, because it is also a function of only $Pr(E|H)$ and $Pr(E)$ and it is an increasing function of $Pr(E|H)$ and decreasing function of $Pr(E)$:

$$E_{SS} = \frac{Pr(H|E) - Pr(H|\neg E)}{Pr(H|E) + Pr(H|\neg E)}.$$

In any separate instance of abduction problem the probability of the evidence $Pr(E)$ is constant, because all of the competing hypotheses have to explain the very same abductive trigger. Both McGrew's and Schupbach and Sprenger's measures of explanatory power are functions of only $Pr(E|H)$ and $Pr(E)$. Hence, if in any instance of abduction $Pr(E)$ does not change its value, then in any instance of abduction problem both measures can be said to be functions only of sole likelihood $Pr(E|H)$.

However, highest likelihood is not sufficient to discern the best explanation. Firstly, as was mentioned before, a hypothesis can entail its evidence and still not explain it. As Lipton (2001b: 110) notes, this is shown by some of the counter-examples to the Deductive-Nomological model of explanation. Secondly, and more importantly,

any association of explanatory goodness with highest likelihood is not sensitive enough to the peculiarities of IBE. For example, if there are i hypotheses that entail the abductive trigger then their likelihoods are equal $Pr(E|H_i) = 1$. On the other hand, different hypotheses can differ according to the degree of unification, simplicity or depth they provide. For example, intuitively a hypothesis that explains five different kinds of phenomena is more explanatory powerful than a hypothesis that explains only one kind of fact even though both hypotheses would entail the phenomena they explain. Hence the explanatory power of hypotheses with the same likelihoods might not be equal.

Prior probability of a hypothesis, $Pr(H)$, is sometimes referred to as the plausibility of the hypothesis and plausibility is sometimes taken to indicate the explanatory power (maximum prior probability account, MPRIOR). More particularly, prior probability together with the likelihood are often claimed to constitute the explanatory power (ML & MPRIOR). According to van Fraassen, criteria of explanatory power that are based on explanatory virtues are rather vague. The precise criteria of explanatory power come from statistical theory, where

> H_1 is a better explanation than H_2 (ceteris paribus) of E, provided:
> (a) $Pr(H_1) > Pr(H_2)$ – H_1 has higher probability than H_2
> (b) $Pr(E|H_1) > Pr(E|H_2)$ – H_1 bestows higher probability on E than H_2 does. (van Fraassen 1980: 22)

Okasha concurs that explanatory power is a composite of prior probability and likelihood:

> The correct way of representing IBE, I suggest, views the goodness of explanation of a hypothesis *vis-á-vis* a piece of

> data as reflected in the prior probability of the hypothesis P(H), and the probability of the data given the hypothesis P(e/H). (Okasha 2000: 703)

Glass (2007: 281) calls the conjunction of ML and MPRIOR the conservative Bayesian approach, according to which H_1 is a better explanation than H_2 if and only if $P(E|H_1) > P(E|H_2)$ and $P(H_1) > P(H_2)$. He calls this criterion conservative, because it cannot provide a total ordering of explanations. Glass claims that ML & MPRIOR is too conservative, because it can fail to rank the hypothesis H_1 as the better explanation than H_2 even if H_1 will provide higher likelihood, $Pr(E|H_1)$, and will have higher posterior probability, $Pr(H_1|E)$, than H_2. This approach can fail to show that the best explanation is also the most probable one, i.e., one with the highest posterior probability. It happens in all cases when H_1 and H_2 have equal prior probability and can happen in many cases when H_1 has lower prior probability than H_2.

Glass (2007) and Niiniluoto (2004) consider posterior probability, $Pr(H|E)$, as a possible criterion of explanatory power (maximum posterior probability account, MPOST). According to Glass, this approach

> seems like the ideal account of best explanation for a defender of IBE. By selecting the best explanation, the most probable explanation is automatically selected. (Glass 2007: 280)

However, MPOST as a measure of explanatory power trivializes IBE. It would identify the most probable explanation as the best one, hence the best explanation would trivially be the most probable (Glass 2010). Moreover, a hypothesis can have the highest posterior probability even if the evidence is unlikely given the hypothesis,

but its prior probability is very high. Therefore, Glass proposes an explanation ranking condition, according to which, if there is any account of the best explanation to replace accounts based on the highest likelihood and highest posterior probability, it should give the same results as the accounts based on the highest likelihood and highest posterior probability whenever these two approaches give the same result as each other (ML & MPOST):

> For two explanations, H_1 and H_2, of a piece of evidence E, if $Pr(E|H_1) > Pr(E|H_2)$ and $Pr(H_1|E) > Pr(H_2|E)$ then H_1 is a better explanation of E than H_2. (Glass 2007: 281–282)

Likelihood still plays the major role in this explanation ranking condition. Posterior probability, $Pr(H|E)$, is the function of prior probability, $Pr(H)$, likelihood, $Pr(E|H)$, and the probability of evidence, $P(E)$. The probability of evidence $Pr(E)$ is constant among the competing hypotheses, i.e., each competing hypothesis has to explain the same abductive trigger. Hence, the ranking according to the explanation ranking condition depends only on likelihoods and priors. If the priors of competing hypotheses are equal or the best explanation has lower prior probability than competitors then the goodness of the best explanation comes from the sufficiently high likelihood. If the best explanation has higher prior probability then the goodness comes from likelihood and prior probability. The introduction of posterior probability serves as its only role to constrain that prior probability should not be too small. Sufficiently high likelihood can offset the prior probability and be alone sufficient to rank a hypothesis as the best one. Moreover, if a hypothesis entails the evidence, then ML & MPOST converts into MPRIOR.

Glass (2007) proposes to determine the best explanation as the

one that best coheres with the evidence. According to him, a satisfactory account of the best explanation should satisfy the explanation ranking condition. The explanation ranking condition is a special instance of the Bovens-Olsson condition, a minimal sufficient condition for the relation "... more coherent than ...":

> for an information pair A, B and probability distributions P and P', if $P(A|B) > P'(A|B)$ and $P(B|A) > P'(B|A)$, then A, B is more coherent on probability distribution P than on probability distribution P'. (Bovens and Olsson 2000: 688)

There are several measures of coherence in the literature, but only one of those measures (Olsson 2002; Glass 2002) satisfies the Bovens-Olsson condition and, consequently, the explanation ranking condition:

$$C_O(A, B) = \frac{Pr(A \wedge B)}{Pr(A \vee B)}.$$

Therefore, according to Glass, one should use exactly this measure of coherence to measure explanatory power. By using Bayes' theorem and assuming that $Pr(A \wedge B) \neq 0$, it can be rewritten as,

$$E_G = C(H, E) = [\frac{1}{Pr(H|E)} + \frac{1}{Pr(E|H)} - 1]^{-1}.$$

Glass calls it the overlap measure of coherence. The best explanation, according to this measure, is a hypothesis that is the most coherent with the evidence, i.e., has the highest value of $C(H, E)$. Nevertheless, as it is evident from the last formula, this measure is still an instance of ML & MPOST.

Niiniluoto (2004: 73) notes that an account, according to which, the best explanation is a hypothesis that maximizes the difference $Pr(H|E) - Pr(H)$, is one more possible probabilistic account of explanatory power. This account is also a version of ML & MPOST

and it also converts into MPRIOR when the hypothesis entails the evidence. However, Niiniluoto does not stick exclusively to only one probabilistic account. For him, ML, MPOST and ML & MPOST are all equally plausible explications of explanatory goodness.

To sum up, probabilistic accounts of explanatory goodness are different combinations of ML, MPRIOR and MPOST. MPOST is a function of likelihood, prior probability (it is also a function of the probability of evidence, but all the competing hypotheses have to explain the same evidence, hence it can be held constant), therefore all the probabilistic accounts are functions of either likelihood $Pr(E|H)$, prior probability $Pr(H)$ or a particular combination of the two. Moreover, if probabilistic explication were an adequate explication of the concept of explanatory power, we do not need to be very precise about the exact relation between ML and MPRIOR in this explication. All other things being equal, we would want a hypothesis with a higher likelihood of having more explanatory power or, all other things being equal, we want a hypothesis with a higher prior probability of having more explanatory power. IBE does not need a cardinal ordering of explanatory power. An ordinal ordering is sufficient for the task at hand: to distinguish the best explanation among the competing ones.

Nevertheless, the thesis is not going to provide a probabilistic account of explanatory power. It is going to be argued further that the explanationist account is a more fundamental account of explanatory power than the probabilistic one and because of that, should be given preference.

1.4.2 Explanatory Likeliness is not Explanatory Power

Explanationists argue it is incorrect to equate the best explanation with probabilistic considerations, and they are correct. There are

several arguments in support of this claim.

Firstly, and most importantly, IBE is an ampliative way of reasoning and Bayesianism is not. Every abductive inference claims more than deductively follows from the abductive trigger, background knowledge and any other data. IBE can suggest new hypotheses. It also suggests how likely one or other hypothesis can be. For example, it claims that a hypothesis that is the best explanation is the most probable hypothesis. On the other hand, Bayesianism is deductive, and therefore it cannot be ampliative. It merely readjusts the probabilities of known hypotheses, it cannot introduce new hypotheses. All the relevant probabilities have to be given before it could determine which of the hypotheses is the best or the most probable. It cannot operate if the relevant probabilities are not given. It cannot suggest how likely one or other hypothesis can be if there are no prior probabilities or likelihoods for them. In other words, deductive inference cannot express an ampliative inference. According to Psillos,

> this content-increasing aspect of IBE is indispensable, if science is seen, at least prima facie, as an activity that purports to extend our knowledge (and our understanding) beyond what is observed by means of the senses. Now, Bayesian reasoning is not ampliative. In fact, it does not have the resources to be ampliative. All is concerned with is maintaining synchronic consistency in a belief corpus and (for some Bayesians, at least) achieving diachronic consistency too. (Psillos 2004: 88–89)

Okasha (2000) and Weisberg (2009) point out that here lies the main advantage of IBE over Bayesianism. According to Okasha,

> in those cases where agents respond to new evidence by

> inventing new hypotheses, the Bayesian model is silent. (Okasha 2000: 707)

Moreover, when prior probabilities do not exist,

> one of the chief advantages IBE has over Subjective Conditionalization is that it provides some basis for preferring one theory over another in such cases. (Weisberg 2009: 133)

The explanationist account of IBE is thus more sensitive than probabilistic. Explanationists would suggest a choice among hypotheses in some cases when probabilistic measures would not be able to determine any difference at all.

Secondly, there are hypotheses that are explanatorily preferable, but are not the most probable. For example, scientists prefer hypotheses that have a greater content (are more informative, are logically stronger), however, a hypothesis with a greater content has more ways of being false than one with a smaller content (Lipton 2004: 116), hence it has to be less probable. Glass (2007: 293) gives an example. A hypothesis H_1 that the die is biased to 2 is more informative than a hypothesis H_2 that the die is biased to an even number. However, H_1 entails H_2 and therefore H_1 simply cannot be more probable than H_2 even though it is more interesting because it makes a more specific prediction. Similarly, consider two nested models, a linear one $y = a+bx$ (LIN) and a parabola $y = a+bx+cx^2$ (PAR). (LIN) is a subset of (PAR), because (PAR) can express (LIN), but not vice versa. According to Forster and Sober (1994: 22), no matter what the likelihoods are, there are no possible prior probabilities consistent with probability theory that would change the fact that $Pr(\text{PAR}|\text{Data}) \geq Pr(\text{LIN}|\text{Data})$. However, scientists sometimes prefer (LIN) over (PAR). Logically stronger or more in-

formative hypotheses are usually more explanatorily powerful than logically weaker or less informative rivals. However, the former, according to the probability theory, cannot be more probable than the latter.

The point that probabilistic considerations might not lead to the most preferable explanation is also illustrated by the base rate fallacy. According to Psillos,

> the base rate fallacy (no matter how one reads it) shows that it is incorrect to just equate the best explanation of the evidence with the hypothesis that has the highest likelihood. (Psillos 2004: 86)

This is because a hypothesis that has a very high likelihood might not have a high enough posterior probability if its prior probability is too low. In this case, if the best explanation were equated with the highest likelihood, the best explanation might not appear to be the most likely one.

Thirdly, explanationists often claim that explanatory power can help to distinguish empirically equivalent hypotheses. Hence, explanatory considerations should go further than empirical considerations. Probabilities, on the other hand, are most often established and adjusted empirically by means of conditionalization. Therefore, if probabilistic criteria are empirical criteria, then the concept of explanatory power has to be broader than mere probabilistic considerations.

Finally, if one equates the best explanation with the probabilistic considerations, then the concept of IBE becomes redundant. Either there is something more than mere probabilistic considerations in explanatory considerations or IBE stands for the same thing as probabilistic considerations and therefore is redundant. If probabilistic

measures of explanatory power were adequate, then IBE would be nothing more than an instance of Bayesianism. IBE thus is a genuine philosophical idea only if there is more in reasoning, especially scientific, than mere Bayesianism can express. Lipton (2004) emphasizes a distinction between the likely and the lovely explanations. The likeliest explanation is a hypothesis that is determined as the most probable on the Bayesian grounds. The loveliest explanation is a hypothesis that would, if true, provide the deepest understanding, and that would be likely because of its explanatory and understanding-enhancing merits. Explanationists bind explanatory power with loveliness, rather than with likeliness. With the latter

> IBE loses all of its excitement (Psillos 2002: 617)

and it

> would in my view take away almost all the interest in the explanationist programme. (Lipton 2001a: 94)

The identification of the best explanation with the most probable explanation would take away all the excitement precisely because

> IBE is an advance only if it reveals more about inference than that it is often inference to the likeliest cause. (Lipton 2004: 60–61)

Explanationists claim that explanatory considerations rather than probabilistic ones determine what it is for an explanation to be a good one. The allegedly exciting idea behind IBE is that the best explanation identified by the explanatory considerations would be the one that is also the most probable one. Therefore, according to Lipton, any proper explication and defense of IBE has to show that

> loveliness is a guide to likeliness. (Lipton 2001a: 94)

On the other hand, Bayesianism can use the Dutch Books argument to argue that the explanationist account does not provide an adequate account of explanatory power and that a probabilistic account should be chosen instead of it. Bayesianism claims that a person is rational only if the probabilities with which she holds her beliefs conform to the axioms of the probability theory and, when she changes the probabilities of her beliefs, the changes conform to the Bayes' theorem. The Dutch Book argument is applied to justify the axioms of the probability theory by showing why the probabilities, with which one holds her beliefs, ought to conform to the axioms. If the probabilities of beliefs did not conform to the axioms of the probability theory, then a person would be susceptible to agreeing to participate in a Dutch Book: a set of bets that would guarantee a sure loss for the person, no matter what the actual outcome of the bets is. There is a distinction between a synchronic Dutch Book—a set of bets that one would accept all the bets at one time—and a diachronic Dutch Book—a set of bets that one would accept over time in some particular order. No matter what the actual set up of a Dutch Book is, it guarantees a sure loss for a participant because of its logical form. Ramsey (1931) and de Finetti (1964) in an article published in 1937 have used a synchronic Dutch Book argument to justify the probability theory. The diachronic Dutch Book version is credited to David Lewis (Teller 1973: 222).

The Dutch Book argument is used by van Fraassen (1989: 160-170) to argue that the only rational way to modify beliefs is the Bayesian one and that if one modifies one's beliefs by preferring best explanations, then these beliefs would become incoherent. Van Fraassen presents an example, where one such person would suffer a sure loss in a Dutch Book. The example analyzes die tossing and probabilities of various hypotheses that the outcome would be

an ace. The die is tossed multiple times and the probabilities of hypotheses are updated according to Bayes' theorem. Then van Fraassen presents us to Peter who updates his beliefs about the ace as an outcome by following a probabilistic version of IBE. According to this rule, Peter first updates his beliefs according to the Bayes' theorem and then slightly raises the probability of the best explaining theory as a bonus for being the best explanation. By modifying his beliefs this way—by adding probability bonuses to the best explaining hypotheses—Peter would think that the set of bets suggested to him is fair, even though he would surely suffer a loss if he accepted the gamble. Assent to separate bets, but not to the bets as a set, shows irrationality. The use of IBE as a rule to modify one's beliefs, according to van Fraassen, is irrational:

> Someone who comes to hold a belief because he found it explanatory, is not thereby irrational. He becomes irrational, however, if he adopts it as a rule to do so, and even more if he regards us as rationally compelled by it. (van Fraassen 1989: 142)

However, van Fraassen's argument is not without flaws. The main idea of the criticism is that van Fraassen's understanding and explication of IBE is incorrect. The probabilistic version of IBE in van Fraassen's argument is not and cannot be IBE. Firstly, in one of the parts of his book titled "What IBE Really Is" van Fraassen argues that IBE is

> a rule to form warranted new beliefs on the basis of the evidence, the evidence alone, in a purely objective manner. [...] there explanation again is an objective relation between hypothesis and evidence alone. (van Fraassen 1989: 142)

Hence, van Fraassen explicates IBE as a function that takes evidence as an argument and gives a hypothesis as a result. However, IBE is a function that takes as its arguments evidence *and* background knowledge. Douven (1999: S426) one more time reminds us that IBE is inferable only with the help of background knowledge and, therefore, is not inferable in every possible context. According to Douven, van Fraassen presents exactly such a context, where a proponent of IBE would not agree to apply IBE. There is no background knowledge. There is only an intentional condition that the die is alien, i.e., that there is absolutely no knowledge how it can behave.

Secondly, despite the fact that van Fraassen in the previous quote correctly describes IBE as a rule that form new beliefs, in his Dutch Book argument IBE is used as a rule to adjust the probabilities of beliefs in response to new data rather than as a rule to form new beliefs.

Finally, there is no IBE in van Fraassen's argument because hypotheses are formed to predict the future rather than to explain something. Peter (in the gamble) guesses the future: what will the outcome of the die toss be. IBE, on the other hand, considers explanations of past events. If there is no abductive trigger, i.e., a fact that would ask for an explanation, then there cannot be abduction, and, consequently, there cannot be IBE.

IBE is susceptible to the Dutch Book argument only if IBE is defined as adding bonus probability to the posterior probability of the best explaining hypothesis. According to Okasha,

> if we accept van Fraassen's way of modeling IBE within the Bayesian framework—as a rule for adding bonus points to explanatory hypotheses—the conflict between IBE and Bayesian rationality constraints is inescapable. (Okasha

2000: 702)

However, explanationists choose a rather different way of incorporating explanatory considerations into probabilistic framework. Explanatory considerations are claimed to determine the very probability distributions rather than simply add explanatory bonus to already determined probability distributions. This is one more argument to show that van Fraassen's explication of IBE is not correct.

To sum up, explanationists do not deny that loveliness correlates with probability. However, they claim, it is explanatory loveliness that determines likeliness rather than likeliness determining loveliness. Explanatory power is seen as probability enhancing and confirmation conducive. High probability and empirical confirmation are claimed to be the by-products of explanatory power. Therefore,

> the goal for defenders of IBE is to give an account of the 'best explanation' in terms of loveliness and show that a feature of such an explanation will be its likeliness, i.e., high posterior probability. (Glass 2007: 280)

This task can be interpreted in two ways: literally and more liberally. Literal reading suggests that if IBE is an inference to the truth of a hypothesis, then after the evaluation of explanatory power the prior probability of the hypothesis has to be set to $Pr(H) = 1$ (or at least close to 1). If IBE is really reliable, then after the conditionalizing and testing of the hypothesis the probability has to remain constant, i.e., it has to remain $Pr(H|E) = 1$ (or at least close to 1). A more liberal reading suggests that after the evaluation of explanatory power the best explaining hypothesis has to be assigned the highest prior probability among the competing hypotheses. If IBE is really reliable, then after the conditionalizing and testing of all the competing hypotheses the best explanation (the one with the highest

posterior probability) has to remain to be the one with the highest prior probability among the competing hypotheses. These readings show the epistemic rationalism in IBE. The mere contemplation of explanatory loveliness should indicate what is true or would be empirically best confirmed. Bayesianism describes how probabilities of beliefs should be changed and updated. IBE, supposedly, indicates what beliefs are true. Hence, it follows that beliefs inferred by IBE should not change their probabilities.

The literal reading corresponds to the deductive aspirations of IBE. The more liberal reading is the reading that shows how explanatory considerations can facilitate the determination of probability distribution. Chapter 2 is going to argue that neither reading is correct, nevertheless, the next section is going to show that the second reading, albeit false, is at least useful if there are no better means of evaluating probability distributions.

1.4.3 Explanatory Power and Determination of Probability Distribution

The section is going to argue that the explanationist explication of the explanatory power is more fundamental than the probabilistic ones, because probabilistic accounts have hardly anything to say about how the probabilities that are used in the evaluation of explanatory power are determined in the first place. The explanationist account, on the other hand, proposes to determine probability distributions on explanatory grounds: to assign a higher marginal probability and likelihood to hypotheses that satisfy explanatory virtues better and vice versa.

If the best explanations were identified with ML or MPRIOR, or their combination, then the best explanations would appear to be the most probable ones after all (the one with the highest posterior

probability). However, it would be trivially so, because explanatory power would be a function of the same arguments as the posterior probability is: both explanatory goodness and posterior probability would be functions of likelihood and prior probability (the probability of evidence is the same among all of the competing hypotheses, therefore it would not influence the final functional comparative value).

Moreover, the identification of the best explanation with ML or MPRIOR faces a further problem: how likelihoods and probabilities are established in the first place. For example, subjective and objective Bayesianists do not agree how the relevant probabilities should be determined. In other words, the best explanation may be the most probable, but then it is not clear what "the most probable" stands for. The identification of explanatory goodness with probabilistic values would require additional means to determine likelihoods and prior probabilities. On the other hand, if there was an alternative way of measuring explanatory power, then it would also be an indirect way to measure the degrees of likelihoods and prior probabilities. Explanationists deal with both of these problems by claiming that explanatory considerations do not simply lead to probability, but that they determine or can be used to determine likelihoods and prior probabilities. This ability is seen as a reconciliation between IBE and Bayesianism.

Bayes' theorem refers only to prior probability and likelihood (the posterior probability is derivative and the probability of the evidence is not about the evaluated hypotheses (moreover, the abductive trigger is a known true fact, therefore $Pr(E) = 1$) or at least it is constant among the competing hypotheses), hence a reconciliation of IBE and Bayesianism through likelihoods and priors is the only way to make the reconciliation. Sober (2002) illustrates the point.

Bayes' theorem implies the following comparative principle:

$$Pr(H_1|E) > Pr(H_2|E) \text{ if and only if}$$

$$Pr(E|H_1) \times Pr(H_1) > Pr(E|H_2) \times Pr(H_2).$$

Hence a hypothesis H_1 can have a higher posterior probability than H_2 if and only if H_1 has a higher likelihood than H_2, a higher prior probability, or both. This in turn implies that

> if 'more plausible' is interpreted to mean higher posterior probability, then there are just two ingredients that Bayesianism gets to use in explaining what makes one hypothesis more plausible than another. This means that if simplicity does influence plausibility, it must do so via the prior probabilities or via the likelihoods. If the relevance of simplicity cannot be accommodated in one of these two ways, then either simplicity is epistemically irrelevant or (strong) Bayesianism is mistaken. (Sober 2002: 22)

If we substitute the talk about simplicity for talk about explanatory considerations, then we get that if explanatory considerations do influence the posterior probability, they must do so via the prior probabilities or vie the likelihoods.

Harman (1970; 1999), Salmon (1970; 1990) and Weisberg (2009) claim that prior probability may depend or should depend on explanatory considerations. Salmon actually uses the term "plausibility considerations" rather than "explanatory considerations", however by plausibility Salmon (1970: 80) means fit with background knowledge (currently accepted scientific theories)—which is coherence. Day and Kincaid (1994), Lipton (2001a; 2004), McGrew (2003), Okasha (2000), Sober (1990; 2001) and Weber (2009) more or less explicitly mention that both prior probability and likelihood

may or should be determined by explanatory considerations. Finally, Niiniluoto (1999b) mentions that explanatory considerations can be manifested in prior and posterior probabilities.

All these suggestions mean that explanatory considerations may, should, or actually determine prior probability distribution and/or likelihood distribution. The better the explanation, the higher its prior probability, likelihood, or both will be. The worse the explanation, the higher probability and/or likelihood of its competitors will be. Most often it is mentioned that the fit with background knowledge affects the plausibility of a hypothesis and, consequently, the relevant probability distributions. However, explanatory considerations include more different explanatory virtues than mere coherence with background knowledge.

Some philosophers claim that explanatory considerations should provide normative constraints for assigning probability distributions within Bayesianism. For example, Psillos (2004: 87) claims that mere permission or possibility to use explanatory considerations in evaluation of probabilities is not an exciting option for the theory of IBE to take. According to him, the more preferable way, if IBE were to be incorporated into Bayesianism, is to take explanatory considerations as normative constraints on the specification of prior probabilities. However, Psillos remains rather skeptical whether this should actually be accomplished. Weisberg (2009: 137), similarly, suggests that explanatory considerations can help the objective Bayesianism to fix objective prior probabilities either by complementing existing objectivist principles or by replacing them altogether. The details of the concept of explanatory goodness, according to him, should provide constraints or fix the prior probabilities of beliefs that agents ought to have.

Huemer (2009a) argues for a way that explanatory considerations

can complement existing objective Bayesianism principles and play a role in determining prior probability. Explanatory considerations, according to him, can facilitate a partial solution to the problem of interpretation of the principle of indifference. According to the principle of indifference or the principle of insufficient reason, when there is no evidence favoring any of n mutually exclusive and jointly exhaustive alternative hypotheses, then one should assign each of them probability $1/n$. However, the principle leads to paradoxes, i.e., contradictory probability distributions. For example, suppose we have a full deck of cards. What is the probability that two cards picked would be both red? According to one possible application of the principle, the probability of picking one red card is $1/2$ and the probability of picking two is the product of single probabilities, i.e., $1/2 \times 1/2 = 1/4$. According to another possible application of the principle, there are three possible outcomes: both cards would be red, both cards would be black, or one card would be red and another one would be black. Hence, the probability of picking two red cards is $1/3$. Both answers are applications of the principle of indifference, but together they are inconsistent, and because of that the principle is judged to be inconsistent. Huemer claims that consistency can be restored if the principle of indifference is applied, i.e., equal probabilities are assigned, at the most explanatory basic level.

> I call this the Explanatory Priority Proviso to the Principle of Indifference. Suppose, that is, that we have two partitions of the space of possibilities, one that divides the possibilities into mutually exclusive, jointly exhaustive alternatives $h_1, \ldots, h_n, \ldots$, and another that divides the possibilities into mutually exclusive, jointly exhaustive alternatives $j_1, \ldots, j_n, \ldots$. Suppose further that each of the

> h_i is explanatorily prior to each of the j_i. Then the former partition should be preferred to the latter for purposes of applying the Principle of Indifference. (Huemer 2009a: 11)

A is explanatory prior to B if A is temporary prior to B, if A is a cause of B, if A is a constituting part of B, if B hold in virtue of A, if B supervenes on A, etc. (Huemer 2009a: 8–9). In the case of our example, a card is a constituting part of a collection of cards, therefore assigning probabilities to individual cards rather than to collections of cards is explanatory prior and, consequently, the first of examples above would be the proper, according to Huemer, way of applying the principle of indifference.

There is also one more way to make reconciliation between IBE and Bayesianism if one wants the best explanations to have the highest posterior probability. One can simply add bonus probability to the posterior probability of the best explanation. This is how van Fraassen (1989) explicates IBE in the framework of Bayesianism. He also shows that this leads to the susceptibility to the Dutch Book and, therefore, incoherence. The Dutch Book argument will be discussed further in the text. For now it suffices to state that if one wants the best explanation to have the highest posterior probability, she should not add bonus probability to posterior probability, but the very prior probability distribution and likelihood distribution should be determined by explanatory considerations. This is probably the only way to escape the menace of incoherence.

Explanationist account seems to be explanatorily prior to the probabilistic accounts. Probabilistic accounts can evaluate competing explanations, but they are of no use if there are no probability distributions for these explanations at hand. On the other hand, the explanationist account can not only evaluate competing explanations, it also provides or at least suggests the means to assess

probability distributions. Hence, probabilistic accounts are neither necessary nor sufficient for the task at hand. The explanationist account is not only sufficient, but can also be considered necessary while there is no better means to evaluate the initial plausibility of explanations.

Chapter 2

Is IBE Truth-conducive?

2.1 Deductive Aspirations of IBE

The conclusion of a deductive inference explicates what follows from its premises. If an inference is deductive, then its conclusion cannot be less true than its premises. If the premises are true, then necessarily the conclusion is true.

IBE aspires to be similar to deduction. A form of inference is deductively valid, if true premises necessarily entail true conclusion. Therefore, IBE would be valid in this sense, if true premises of IBE would necessarily entail that the conclusion of IBE is true. The premises of IBE have the following form:

> The surprising fact, C, is observed;
> But if A were true, C would be a matter of course,
> No other hypothesis can explain C as well as A does.

The conclusion of IBE has the following form:

> Hence, A is true.

Therefore, IBE would be valid, if the capacity of a hypothesis to explain the abductive trigger together with the hypothesis being the best explanation for the abductive trigger would entail that the

hypothesis is true. This proposition is exactly what the theories of IBE argue for: an explanatory power of a particular (highest among the competitors and sufficiently high in itself, e.g., better than its negation) degree is a sufficient condition for the truth of an explanatory hypothesis. In other words, the theories of IBE claim that if a hypothesis actually explains some phenomena of interest and if there is actually no better explanation than the hypothesis under consideration, then it is impossible that the hypothesis is not true or at least is not close to truth.

On the other hand, this validity depends not on the logical form of the inference, as is the case of deductive inference, but on the particular propositional content of the particular inference: logical, ontological, nomological, etc. relations contained in the background knowledge and hypothesized by the explanatory hypothesis. It means that IBE is material inference, not formal. Deductive forms of inference, given true premises, have true conclusions because of their formal features: the conformity of their logical forms to the logical truths. IBE, given true premises, can have true conclusions only because if the ontological, nomological, etc. relations (in addition to logical ones) it makes us of are true.

IBE resembles deduction in several other aspects. The strongest resemblance between IBE and deductive inference lies in the fact that the logical form of the evaluation of the explanatory power among the competing hypotheses in IBE mimics the disjunctive syllogism. Firstly, one selects a set of potential explanations: this forms a disjunction of explanatory hypotheses. Then, one reviews and rejects hypotheses from this disjunction due to their insufficient or inferior explanatory power till only one most explanatory powerful hypothesis is left, which comes to be declared true.

Secondly, deductive conclusions cannot be less true than their

premises. In other words, deductive conclusions are necessarily equally true or more true than their premises. This is most evident in the multivalued logic where the truth value of a conditional (and, analogously, of an inference) is calculated by the equation

$$\psi \to \varphi = \min(1, 1 - \psi + \varphi)$$

or, in other words, where the conditional is true, $\psi \to \varphi = 1$, if and only if the truth value of the antecedent is less or equal to the truth value of the consequent, $\psi \leq \varphi$. A similar feature is ascribed to IBE. Proponents of IBE claim that the conclusions of IBE are often more probable than their premises. For example,

> because of [...] and the unifying power of the best explanation, we may have more reason to believe a theory than we initially had for believing the data that supports it (Lipton 2004: 205)

or

> abductions often display emergent certainty; that is, the conclusion of an abduction can have, and be deserving of, more certainty than any of its premises. (Josephson and Josephson 2003: 15)

IBE performs a comparative evaluation—evaluating which of the competing explanations is the best one—but infers an absolute conclusion—the best explanation is true (not simply that it is most likely to be true among the competing explanations). However, only deductive forms of inference allow us to infer absolute conclusions. Thus appear the deductive aspirations of IBE: even though IBE is merely a form of inductive inference and can provide only a probable conclusion, IBE is claimed to produce an absolute conclusion.

Proponents of IBE are aware that IBE is not and cannot be deductively valid. Nevertheless, theories of IBE maintain that IBE is truth-conducive, that IBE is a reliable rule of inference, and that it provides a good strategy to look for truths. The present chapter is going to evaluate these deductive inspirations: what are the reasons to believe in the truth of a hypothesis given that it is the best explanation for some abductive trigger of interest?

2.2 Reliabilist-Coherentist Justification

Reliabilism and coherentism constitute the main strategy that proponents of IBE use to epistemically justify IBE. More particularly, proponents of IBE appeal to the reliability of IBE and justify the reliability with an appeal to coherentism.

Process reliabilism claims that a belief is justified in case the belief is a product of a reliable belief formation process, i.e., a process that leads to a high proportion of true beliefs or that is prone to produce true beliefs rather than false ones. Thus a conclusion of IBE is claimed to be justified, because it is the product of IBE and IBE is a reliable belief formation process. As Lipton puts it,

> I take it that our inductive practices are reasonably reliable, certainly better than random guessing. (Lipton 2004: 145)

The reason for this confidence in its reliability is claimed to be several-fold. Firstly, the past performance of IBE is an indicator of future performance. The past performance was successful and an inductive generalization is made that future performance will also be successful (Lipton 2007: 459–460, Psillos 1999: 81–91). Secondly, the past performance of IBE is an indicator of what features of IBE were responsible for the successes and failures and this allows us

to modify the use of IBE so as to improve its performance (Lipton 2007: 460). Thirdly, the innateness of IBE as a human cognitive capacity shows that this capacity is reliable (Carruthers 1992: 110). Fourthly, the constantly accumulated knowledge constrains and refines the set of potential explanations and the set of reliable and useful explanatory considerations thus making IBE even more reliable as times passes. This process is reliable, because it safeguards the coherence of the total belief corpus (Lipton 2004: 148; 2007: 460; Psillos 2002: 619).

The third reason is the most important and is considered to be the most powerful justification of the reliability. This is the part where coherentism is used to justify IBE. Coherentism claims that a belief is justified if the belief coheres with other beliefs that are or would be believed. This condition is in some respect equivalent to the definition of IBE. IBE is an inference to the explanatory hypothesis that best coheres with background knowledge. Hence description is both a definition and a justification of IBE.

Lipton (2004: 148–151) describes the process of IBE as an application of the twofold filter of explanatory loveliness on the set of all the possible explanatory hypotheses (our division of IBE into an abductive step and an evaluative step corresponds to this twofold filter). This is the paradigmatic example of the use of coherence in IBE. In the first filter, background knowledge favors those possible hypotheses that cohere with it and thus discriminates a shortlist of potential explanatory hypotheses. In the second filter, background knowledge determines the set of relevant explanatory considerations and applies them to select the actual (the best and, supposedly, true) explanation out of the set of earlier chosen potential explanations. The background knowledge is considered to be true and this truth-likeness is claimed to transfer to the conclusion of IBE. Moreover, it

is not the case that the background knowledge itself is taken as not requiring any justification. Its only justification is that it is itself the product of prior applications of IBE. Background knowledge determines what is lovely and loveliness reciprocally determines what is going to become future background knowledge. Hence, coherence with background knowledge directs the selection of potential explanations, the evaluation of potential explanations and determination of the best explanation. All this ensures that no belief or proposition will become a part of accepted knowledge, if it does not cohere with what is already accepted as true. Similarly, Psillos maintains that explanatory coherence is

> a vehicle through which an inference is performed and justified. IBE is the mode of inference which effects ampliation via explanation and which licenses conclusions on the basis of considerations which increase explanatory coherence. (Psillos 2002: 619)

IBE is an inductive kind of inference, because every instance of IBE is underdetermined by the evidence. There is always at least a logical possibility of an alternative explanation of the abductive trigger. The set of potential explanations can never be exhaustive and complete. The actual explanation can appear to be the one that was not considered as a potential explanation. Van Fraassen elaborates this line of reasoning and presents an argument known as the bad lot argument or the argument from underconsideration. IBE selects the best from the already given pool of explanatory hypotheses. It is impossible to know whether we have considered all possible potential explanations. Hypotheses that are not proposed and are not known are not evaluated. The actual explanation can be among those unknown hypotheses. Hence,

> our selection may well be the best of a bad lot. (van Fraassen 1989: 142)

If the theories of IBE, nevertheless, insist that one can still make an absolute inference to the truth of an explanatory hypothesis from a comparative evaluation, then, van Fraassen argues, the proponents presuppose a privileged access to the truth:

> for me to take it that the best of set X will be more likely to be true than not, requires a prior belief that the truth is already more likely to be found in X, than not. (van Fraassen 1989: 142)

If the theories of IBE do not presuppose privileged access, then it is very unlikely that the actual true explanation will be among known hypotheses rather than among many hypotheses that would explain the abductive trigger, but are not yet formulated and maybe never will be.

Peirce indeed argues that people have a privileged access to truth:

> man has a certain Insight, not strong enough to be oftener right than wrong, but strong enough not to be overwhelmingly more often wrong than right, into the Thirdnesses, the general elements, of Nature. An Insight, I call it, because it is to be referred to the same general class of operations to which Perceptive Judgments belong. This Faculty is at the same time of the general nature of Instinct, resembling the instincts of the animals in its so far surpassing the general powers of our reason and for its directing us as if we were in possession of facts that are entirely beyond the reach of our senses. It resembles instinct too in its small liability to error; for though it goes wrong oftener than right, yet the relative frequency with which it is right is

> on the whole the most wonderful thing in our constitution. (Peirce 1934: 5.173)

More particularly, Peirce (1958: 7.220) argues that abduction would make no sense if people did not have an instinct "in a finite number of guesses" to stumble upon the correct hypothesis. Nevertheless, abduction is successful and because of that one is forced to infer that people do indeed possess this kind of instinct. Hence, Peirce argues for the privileged access to truth in regard to abduction. Abduction provides potential explanations for IBE to choose among. Therefore, if Peirce is right, the privileged access to truth should hold for IBE as well: the truth, i.e., the actual explanation, will be among the finitely many guesses selected for the evaluation of explanatory power.

For Lipton the privilege is guaranteed by the reliability of IBE. More particularly, Lipton (2004: 157–159) argues that the bad lot argument rests on two premises that are mutually incompatible. One of these premises states that there is no privileged access to the truth (the no-privilege premise), i.e., that there is no reason to believe that the actual explanation will always be among the potential explanations that are chosen for the evaluation of their explanatory power. The other of the premises states that the ranking of explanatory power is reliable (the ranking premise), i.e., a hypothesis that is the closest to the truth will always be evaluated as the best among its competitors, even though one would not be able to know how close to the truth the hypothesis actually is. As mentioned before (1.3.1), background knowledge is the medium in which one evaluates and selects the best explanation. The ranking premise states that the evaluation is reliable. However, for the evaluation to be reliable the background knowledge, which is used for evaluation, also has to be true. If the background knowledge were not true, the evaluation would be distorted and the less true hypotheses would be

often evaluated as better than the more true ones. In other words, if background knowledge were not true, then the ranking premise could not be true. Therefore, if one accepts the ranking premise, one cannot reject the truth of background knowledge. Background knowledge is the product of previous instances of IBE and current instances of IBE will become a part of future background knowledge. If the ranking premise is true, then best explanations have to be true rather than merely closer to the truth than their competitors. But this is possible only if the actual explanation is always in the set of potential explanations chosen for the evaluation, i.e., only if the no-privilege premise is false. Therefore, according to Lipton, if the ranking premise is true, then the no-privilege premise has to be false, the whole bad lot argument self-destructs and the truth claims of IBE are safeguarded.

Lipton's reply to the bad lot argument also rests on two premises: the very same ranking premise and the premise that the background knowledge is true. According to him, these two premises entail that there always be the actual explanation among the generated hypotheses. However, the ranking premise and the truth of background knowledge are also the claims—the truth of which—the theories of IBE want to establish in the first place. The bad lot argument may self-destruct in Lipton's counter-argument. However, this does not prevent the doubt that the premises of Lipton's argument might not actually hold in the first place.

The presence of background knowledge is very important to counter the bad lot argument. Psillos (1999: 217) concurs with Lipton that the bad lot argument works only on the assumption that theory choice operates in knowledge vacuum. However, the background knowledge might not be true. The pessimistic induction argument (2.5.2) shows that it was often false. The major reason to believe in

the truth of background knowledge is IBE itself. The other reason is a blind belief in its truth. As Psillos puts it,

> undeniably, realists take an extra epistemic risk when they say that background theories are approximately true; but taking an extra risk is the necessary consequence of aspiring to push back the frontiers of ignorance and to get to know more things. (Psillos 1999: 222)

In this quote the truth of background knowledge is only a presupposition and its only justification is pragmatic: we do not know whether background knowledge is true, but if we accept it tentatively as true, then it can help us accomplish something else.

The possibility of conflicting ordering of explanatory loveliness suggests that the ranking of explanatory loveliness might not be reliable. Ladyman et al. (1997) appeal to the possibility that there might not always be one best explanation:

> that conclusion would require (at least) one further premise, viz., that there is (almost) always a unique best explanation, i.e., that the ordering of explanations for e according to some standard of 'goodness' almost always has a greatest element. But what justification is there for this premise? (Ladyman et al. 1997: 309)

The possibility of conflicting orderings emerges because different explanatory virtues may be considered as possibly mutually incommensurable; therefore there always remains a possibility that in each particular case it is impossible to provide the univocal best explanation. This can happen because of several reasons.

Some explanatory virtues are not universally acceptable or applicable. Simplicity, for example, is one of the most often cited explanatory virtues, however Carruthers notices that

> appeals to simplicity should cut little ice in the biological realm. On the contrary, we should expect biological systems to be messy and complicated, full of exaptations and smart kludges. (Carruthers 2006: 151)

Different definitions of simplicity lead to contrary orderings of explanatory loveliness. Sober (2001) gives an example of two nested models, one is linear $y = a + bx$ (LIN) and another is parabola $y = a+bx+cx^2$ (PAR). (LIN) has two adjustable parameters (a and b) and (PAR) has three (a, b and c), hence if one takes simplicity to stand for fewer adjustable parameters then (LIN) is simpler model. However, if one takes simplicity to stand for fewer assumptions then (PAR) is simpler than (LIN), because (LIN) is equivalent to the conjunction of (PAR) with an additional assumption that $c = 0$. If one defines propositional logic only with the Sheffer stroke one will have a semantically simpler system, but a much more complex system syntactically. Lakatos goes as far as to claim

> No doubt, simplicity can always be defined for any pair of theories T_1 and T_2 in such a way that the simplicity of T_1 is greater than that of T_2. (Lakatos 1970: 131 note 106)

Assessment of unification virtue can also give conflicting orderings. H_1 can have a larger cardinality of its consequence set than H_2, but H_2 can explain more important facts. The MMH measure of unification can rank hypotheses according to the degree of positive relevance they bring, but this measure is incapable ranking hypotheses that entail their evidence.

Finally, different explanatory virtues may provide different conflicting ordering. For example, a very deep and precise explanation would most often not be a very unifying one and vice versa. A very broad explanation is not likely to be very precise.

Schurz (2008) distinguishes 15 kinds of abduction patterns and sub-patterns and claims that all of them employ different evaluation criteria for explanatory power, different criteria can come into mutual conflict and this implies that there is no unique criterion to establish the best explanation. Persson (2007: 142–143) claims that different concepts of explanation may select different hypotheses as potential explanations and these selections, not to mention their orderings of explanatory loveliness, might not even have common members. For example, deductive-nomological, causal mechanical or unificationist accounts of explanation may prefer different hypothesis as the best ones. Even the sole causal mechanical account can produce conflicting orderings if different accounts of causation (e.g., counterfactual and manipuliationist) will favor different hypotheses.

Lipton (2004: 142–144) dubs as 'Hungerford's objection' a pattern of arguments that claim that explanatory loveliness is too subjective (depends on the eye of the evaluator) to give a decisive ordering. He claims that what counts as warranted inference depends on available evidence, background knowledge, notion of explanation, etc. which are themselves relative to the audience, hence, the relativity of explanatory loveliness is no more extreme than any of these components. But this response does not escape the relativity; therefore it does not provide any reason to suppose that there cannot be incommensurable ordering of explanatory loveliness. On the contrary, Lipton's response only states that explanatory loveliness is no more subjective than any other epistemic activity. Hence, it only strengthens the suspicion that the theories of IBE do not escape the menace of conflicting and incommensurable orderings.

The inter-relations between explanatory virtues (1.3.2) reduce the possibility of incommensurability, but it is unlikely that it can be totally escaped. There can be several good explanations that have

their advantages in different aspects of explanatory loveliness. One can reply that only when there is one unambiguous best explanation is one entitled to infer the truth of the explanation. If there is no uniquely best one, then one should not infer to the truth of the allegedly best one. However, what the theories of IBE claim is that if one wants to find a cause, she should look for the best explanation. Moreover, this best explanation does not have to be the uniquely best one, it suffices, as Lipton (2004: 56) states, that it is good enough. One may also respond that a rational discussion can help settle the question of which explanatory virtues, in a particular instance, are the most important. However, this response is a dangerous path to take, because it can easily lead to situations as described in social constructivism studies (e.g., Latour and Woolgar 1986; Collins and Pinch 1993): it can lead to theories that are socially constructed rather than discovered.

Hence the premises of Lipton's own counter-argument to the bad lot are not without flaws. The background knowledge can be false and the reliability of the ranking of explanatory power might not be reliable. Nevertheless, there is another counter-argument to the bad lot: the catch-all hypothesis argument.

The bad lot argument works only if one makes an absolute inference from a comparative argumentation. If one had the exhaustive set of explanatory hypotheses, she would be deductively justified, by disjunctive syllogism, to make an absolute inference from a comparison. The theories of IBE elaborate this line of reasoning and propose a way to acquire an exhaustive set of hypotheses. When there is only one explanatory hypothesis, evaluate the explanatory power of this hypothesis in relation to the explanatory power of its negation. When there are more than one competing explanatory hypothesis, add a catch-all hypothesis that all the competitors are

false (Lipton 2004: 155–157; Minnameier 2004: 88; Niiniluoto 1999b: S447).

It may appear that this move warrants an absolute inference from comparative considerations. The addition of a catch-all hypothesis—a complement that would make the relevant set exhaustive—to the set of abductive conclusions would make IBE an instance of disjunctive syllogism and, consequently, would make IBE deductively valid. However, as Douven (2002) notes, the move is highly dubious. The all-negating catch-all hypothesis would generally be hardly informative. It would be hard to test, because generally its empirical consequences would not be evident. Therefore,

> qua explanation it will be ranked quite low (if it will be ranked at all, which would seem nonsensical in case it is unclear what empirical consequences it has). (Douven 2002: 357)

Simple examples perfectly illustrate Douven's claim. Suppose we observe that the light in a room went out. One potential, and suppose the actual one, explanation for this is

(1) The light switch was turned off.

The corresponding catch-all hypothesis is

(2) The light switch was not turned off.

Another potential explanation can be

(3) The fuse blew out.

The corresponding catch-all hypothesis is

(4) The fuse did not blow out.

The catch-all hypothesis for both (1) and (3) is

(5) The light switch was not turned off and the fuse did not blow out.

There are three possible cases of evaluation of the explanatory loveliness: (1) can be evaluated against (2); (3) against (4); and (1), (3) and (5) among themselves. The true propositions are (1) and (4). The false propositions are (2), (3) and (5). Hence, the catch-all strategy seems to work, because in each of the three cases there will be a true hypothesis among the competing hypotheses.

Nevertheless, let's look closer at the second case where the false (3) is evaluated against the true (4). Even if false, (3) is at first sight a very good explanation. Very often lights go out exactly because fuses blow out. In order to check whether it is actually the case one has simply to go to the fuse and look if it has or has not blown out. On the other hand, even if true, (4) is not a good explanation. Its only empirical consequence is that the lights should keep shining, because the fuse is working perfectly, and this contradicts the abductive trigger. One may argue that (4) is not an explanation at all, because it does not presuppose any mechanism, causal story, nomological laws, etc., i.e., it does not have any depth; it does not provide understanding of why the abductive trigger occurred. As an explanation the proposition "The light in the room went out, because the fuse did not blow out" is an equivalent explanation to the proposition "The light in the room went out, because the sofa in the room did not change its color." If this catch-all hypothesis was a good enough explanation, then any true proposition, albeit totally unrelated, would be as equally good an explanation as the catch-all. Hence, in situations of this kind IBE would force one to choose a good enough explanation that is a false proposition instead of the true proposition that is not a good enough explanation.

Remember that we defined abduction as the first step of IBE

(1.2). A hypothesis is accepted into the set of potential explanations only if it explains the abductive trigger. If a hypothesis does not explain the abductive trigger, it cannot be accepted into the set of potential explanations. Hence, the catch-all hypothesis would not be considered as a potential explanation if it did not explain the abductive trigger, i.e., if it were not a possible conclusion of abduction. However, the catch-all hypothesis is a complement to the set of existing potential hypotheses; therefore it is a product of complementation and not of abduction. The catch-all hypothesis would state that all the other potential explanations are false, that the actual explanation is not among the other potential explanations, so it would claim things that are opposite to the things that are claimed in the other potential explanations, therefore, the catch-all would explain the opposite things to the abductive trigger, i.e., that the opposite to the abductive trigger has to be true. To repeat an example, a blown fuse is an explanation for the failure of the light bulb, but a catch-all for it "The fuse did not blow" will not explain the failure of the light bulb, because, according to this fact, the light bulb should remain lit. Therefore, the catch-all hypothesis cannot be used to save IBE from the bad lot argument. Consequently, the catch-all hypothesis cannot be used to make the set of potential explanations exhaustive, so it cannot be used to save IBE from being deductively invalid. IBE inherits deductive invalidity from abduction and always remains susceptible to the bad lot argument.

One can provide a more explanationist argument against the use of a catch-all hypothesis. A particular instance of IBE is claimed to work only because of particular substantial and contingent assumptions in the background knowledge that cannot be neatly formalized (e.g., Psillos 2002). These substantial assumptions explain why a particular explanation is a lovely explanation. Now, it follows that

a catch-all hypothesis can be evaluated in regard to its explanatory loveliness also only with the help of substantive background assumptions. The assumptions describe how the explanatory entities in the hypothesis bring about the abductive trigger. If there are no explanatory entities, then it remains unknown how the abductive trigger is brought about. A proposition stating that something is not the case, hence, is not an explanation and cannot be an explanation. Its explanatory depth is zero. Consequently, a catch-all hypothesis will never be a lovely explanation if it does not have any positive content. On the other hand, if a catch-all hypothesis does have a positive content, if it posits some explanatory entities, then the bad lot argument would be applicable to it. The bad lot argument would ask what the reason is to believe that these particular explanatory entities do the actual work rather than some as-yet unknown or unthought entities. Therefore, a catch-all hypothesis would not be a good explanation or the bad lot would be applicable to a catch-all hypothesis.

To sum up, Psillos claims that IBE strikes the best balance between ampliation and epistemic warrant. Lipton maintains that reliability of IBE is sufficient to counter the Humean skepticism about the truth claims of inductive inferences. However, these claims are dubious. Firstly, Lipton claims that IBE is no worse in respect to the Humean skepticism than any other inductive method of reasoning. The reply in its form is the same as Lipton's reply to the Hungerford's objection, and it equally badly misses the point. IBE being no worse than other inductive methods of reasoning does not imply that it is any better. Secondly, as we have shown earlier, coherentist justification of the reliability of IBE is not without flaws. The background knowledge is not necessarily true. The ranking of explanatory loveliness might not be unambiguous and consistent. The

actual hypothesis might not necessarily appear among the hypotheses chosen for the assessment of explanatory loveliness. The fact that IBE usually worked in past does not guarantee that it will continue to work in the future. The reliabilist-coherentist justification of IBE too often uses inductive generalizations to establish a point. The conjunction of all these generalizations, some of which are questionable themselves, makes it very unlikely that IBE is truth-conducive.

2.3 Psychological Adequacy, Pragmaticism and Evolutionary Justification

2.3.1 The Psychological Hypothesis

Studies in epistemology and philosophy of science rarely ask whether IBE is typical to humans as such or sometimes even explicitly doubt that people in fact invoke this form of reasoning in their ordinary way of thinking. For example, van Fraassen claims:

> [...] we can have no good evidence for the psychological hypothesis that people do in fact follow the rule of inference to the best explanation. (van Fraassen 1985: 295 fn. 19)

On the other hand, there are philosophers that declare the psychological adequacy of IBE (following van Fraassen we will call this claim the psychological hypothesis). For example, there are philosophers of mind who explicitly endorse that reasoning by IBE is a psychological fact. According to Fodor:

> it appears that much of what the mind does best is 'abduction' or 'inference to the best explanation.' (Fodor 2000: 97)

Carruthers (1992; 2006) also judges IBE to be a distinctively human cognitive capacity. These philosophers do not argue for the psycho-

logical adequacy of IBE, but accept it as a self-evident fact. IBE is interesting for them as an object of study only to the extent that it needs an explanation as to how it originated or how exactly it is implemented in the mind. More particularly, (Carruthers 1992: ch. 7) claims that IBE is an innate capacity, because people possess it even though they are not explicitly taught it, and because it does not seem to be learned from experience. Fodor (2000), on the other hand, tends to argue that abduction (Fodor, as can be seen in the quote above, means by this term the same thing as IBE) is not likely to be explained by any current theory of mind.

The present section enumerates empirical results in support of the psychological adequacy of IBE. This adequacy is relevant for the thesis, because the pragmatic and evolutionary justifications of IBE rest on the truth of the psychological hypothesis.

The idea that IBE is a human cognitive capacity is not recent. Theories of IBE merely develop Peirce's claims about abduction or, if the distinction between early and late Peirce (Gabbay and Woods 2005: 40; Psillos 2009a: 131) is tenable, the claims of late Peirce. On the one hand, abduction for Peirce is a logical inference and the only way to introduce new ideas (Peirce 1932: 2.96). On the other hand, abduction for Peirce is also an instinct to guess the right kind of hypotheses, and the postulation of this instinct is the only way to explain the high rate of successful scientific hypotheses. Even though the instinct is not infallible, it is much more successful than pure chance would allow (Peirce 1934: 5.172–5.173, 5.591; 1958: 7.220). Abduction for Peirce is a form of inference and a cognitive capacity at the same time. Identification of abduction with an instinct makes abduction not just a subject matter of logic, but rather a subject matter of psychology (Paavola 2005: 143).

The logical form of IBE is abductive (1.2) and remains abductive even after the evaluation of explanatory power (2.2). Every instance of IBE is abduction, but not every instance of abduction is IBE. Both abduction and IBE are non-deductive forms of arguments, because the logical form of abduction and, consequently, of IBE is deductively invalid. If someone reasons abductively and accepts the abductive conclusion as true, she makes a logical fallacy known as Affirming the Consequent (AC). Due to its abductive mechanism IBE is deductively invalid and should be considered as an instance of AC as well. However, experiments show that people, nevertheless, maintain AC as a valid form of inference rather often. Knauff (2007) summarized the findings from a number of classical studies that explored whether people do perceive various deductively valid and invalid forms of inference involving a conditional premise as valid or as invalid. The summary of five studies shows that around half of the participants (from 27% to 75%) treated AC as valid (Knauff 2007: 21). This suggests that abduction as the logical form of IBE is often perceived as valid even though it is not actually valid. Some people would accept the conclusions of IBE as true although logic does not permit that.

IBE is often characterized as an inference to the hypothesis, which, if true, would be the best explanation or would provide the most understanding (e.g., Lipton 2004). Studies show that an explanation why a hypothesis can be true raises the perceived probability of that hypothesis. Koehler (1991) in the section of his article "Explaining is Believing" enumerates experiments whose results indicate that an explanation why a possibility may turn out true raises the confidence in the truth of that possibility. First of all, a generation of an explanation of why some future events can occur raises the perceived likelihood of the actual occurrence of these events. For

example, subjects predicted a victory of that college football team whose hypothetical victory they were assigned to explain prior to the prediction. Secondly, creation of an explicit explanation enhances belief perseverance. That is, beliefs for which subjects were asked to provide explanations, persevere and continue to be held true even after the evidential basis for the explanations has been removed or refuted.

Two further studies reveal how an ability to explain is sometimes used as an evidence for belief. A study by Koslowski et al. (2008) shows that people more often accept some information as evidentially relevant in order to explain some event when there is a broader causal explanation that can accommodate this information than when such an explanation is absent. Brem and Rips (2000) show that people tend to use explanations as a substitute for evidence when evidence is missing, insufficient or is difficult to come by. Evidence is required to test hypotheses. Relevant evidence either raises or lowers the probability of a particular hypothesis. These two studies indicate that people sometimes treat an ability to explain in a similar way as evidence when they want to support a claim. Hence, in these experiments people used explanations to raise the probability of particular hypotheses.

Koehler (1991)'s words nicely summarize all those experiments:

> The theme that emerges through the examination of this empirical work is that any task that requires a person to treat a hypothesis as if it were true can strengthen the confidence with which that hypothesis is held. (Koehler 1991: 499)

In other words, people believe in those hypotheses that, if true, would explain some event. According to the theories of IBE, actual

explanations are those that, if true, would provide the best explanation for an event. On the one hand, these results seem to support the psychological adequacy of abduction rather than that of IBE. People accept as true hypotheses that would explain, but not necessarily hypotheses that are the best explanations. However, people take explanations to be true (the feature of IBE) and not merely possibly true (the feature of abductive inference). People seem to be satisfied with even less demanding requirements than IBE asks for. Hence, the results of these experiments tend to support the psychological hypothesis.

Explanatory virtues explicate what it is for a hypothesis to be a better explanation. The most commonly mentioned explanatory virtues are the virtues of coherence, unification and simplicity. Their role is twofold. Firstly, explanatory virtues are claimed to evaluate and rank the explanatory power of competing explanations (1.3.1). Secondly, explanatory virtues are claimed to evaluate prior probabilities and likelihoods (1.4.3). Experiments show that people employ explanatory virtues in both of these ways.

Thagard (1989) in his theory of explanatory coherence states that people accept broader, simpler and deeper explanations as better. Read and Marcus-Newhall (1993) conducted experiments to test different aspects of this theory. They discovered that subjects value narrow explanations as better than broad explanations when explaining singular facts, although broad explanations are judged to be better than narrow explanations when explaining the multiplicity of facts. Breadth in this study is defined as an ability to explain more facts; hence, it is used as a synonym for unification. Next, Read and Marcus-Newhall discovered that in order to explain a multiplicity of facts broad explanations are evaluated as much better than conjunctions of narrow explanations. The authors claim that this

result shows that people prefer simpler explanations. Finally, Read and Marcus-Newhall observed that explanations are perceived to be better when they are explained by a further explanation than when they are not, i.e., deeper explanations are preferred.

Lombrozo (2007) examined only a sole explanatory virtue of simplicity, but her results are very comprehensive and strongly support both claims, distinguished above, that proponents of IBE associate with explanatory virtues. Lombrozo conducted several experiments that tested the relationship between simplicity and probability of explanatory hypotheses. One experiment showed that people preferred simpler explanations when information about their probability was absent and preferred more probable explanations when information about their simplicity was absent. Other experiments showed that people assign higher prior probability to simpler explanations and that complex explanations are valued more highly than simple explanations only after disproportionate evidence for the complex ones is given. Finally, one more experiment showed that only when information about probabilities of explanations is unambiguous do people prefer more complex hypotheses to simpler ones. Lombrozo's main conclusion is that simpler explanations are assigned a higher prior probability when there is no clear information about their probabilities and preference of simpler hypotheses ceases when that information is revealed. These results are in line with the theories of IBE, especially the claim that considerations about simplicity as an explanatory virtue contribute to the assignments of prior probability and the claim that simplicity as an explanatory virtue can trump probability when evaluating hypotheses.

Information about the underlying mechanism is one of the explanatory virtues associated with IBE. Ahn et al. (1995) examined whether people seek information about covariance or about causal

mechanisms when asked to provide an explanation for an event. Experiments showed that people prefer information about underlying causal mechanisms rather than covariance both when asking for further information about the events to be explained and when providing their explanations for these events.

A neuroimaging study by Harris et al. (2008) can also be interpreted as in line with the psychological hypothesis. It revealed that the acceptance of a statement as true is associated with a particular part of the brain (the ventromedial prefrontal cortex) and the rejection of a statement as false is associated with the activation of another particular part of the brain (the anterior insula). The former association means a link between belief and emotion and the latter association means a link between disbelief and the sensation of taste, pain perception and disgust. Harris et al. concluded that the final acceptance or rejection of a statement appears to rely on hedonic processing because it is partially governed by the same regions of the brain that govern hedonic judgments. This result that links belief in a statement with the feeling of pleasure and disbelief with the avoidance of disgust vindicates Lipton's choice of the term "loveliness" to stand for the explanatory goodness of a hypothesis or understanding that the best explanation can provide. Even though the denotation of this word in the context of IBE is strictly epistemic (Barnes 1995: 273 fn. 4), the word as such has rather emotional, aesthetic and hedonistic connotations.

Background knowledge is one of the most important things in discerning the best explanation. Experiments show that background knowledge contributes to the credibility of explanations and that coherence with background knowledge is a condition for a piece of information to be accepted as true.

The study by Koslowski et al. (2008) already mentioned showed that people grasp some evidence as more relevant when it can be incorporated into an explanation. What this study also indicated is that explanations become more credible when they can accommodate relevant background information. In other words, the perceived probability of explanations is a function of their coherence with background knowledge:

> explanations become increasingly convincing as evidence mounts up that connects the explanation in a causal way to what else there is in the world that we have fairly good reason to believe. (Koslowski et al. 2008: 483)

The role of background knowledge is further scrutinized by Richter et al. (2009) whose experiment shows that background knowledge conducts validation of incoming information. They claim that their results indicate the existence of quick and efficient cognitive mechanisms. If background knowledge is accessible, integrated, relevant, and held with a high subjective certainty, these mechanisms accept beliefs that are coherent with background knowledge and reject those beliefs that are not. A neuroimaging study by Marques et al. (2009) supports the findings of Richter et al. Marques et al. discovered that verifying true statements activates the left inferior parietal cortex and the caudate nucleus and conclude that this is a neural correlate compatible with an extended search and matching process for particular stored information. Accordingly, they observed that verifying false statements activates the fronto-polar cortex and conclude that this is consistent with the claim that the processing of false statements involves a search for contradiction between information in statements and information stored in memory. Even though these two studies do not deal directly with explanations, they do it

indirectly, because every explanation is a statement or a set of statements. These studies support the claim that coherence with background knowledge plays a decisive role when evaluating the truth of incoming information.

The famous and often replicated study by Tversky and Kahneman (1982) also shows that background knowledge is relevant for perceived probability. In Tversky and Kahneman's experiment subjects were given a piece of particular background knowledge and had to evaluate the probability of a set of statements. According to the probability theory, a conjunction cannot be more probable than any of its constituents. However, contrary to the requirements of the probability theory, more than 80% of participants evaluated a conjunction as more probable than one of its conjuncts and committed the so-called conjunction fallacy. In other words, given the particular background information (Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.), most people think that the statement "Linda is a bank teller and is active in the feminist movement" is more probable than the statement "Linda is a bank teller", which is impossible, according to the probability theory. This result, however, should be expected if coherence with the background knowledge really influences the perceived probability. Moreover, this result also gives some more credence to the claim that explanatory considerations (coherence in this particular case) contribute to the determination of prior probabilities (Meijs and Douven 2007: 356 fn. 12).

IBE is a form of inference based on comparative evaluation. It allows one to infer the truth of some hypothesis only if there is no better explanation for the phenomena at hand. Experiments

show that availability of competing hypotheses influences perceived probabilities.

Koehler (1991) not only summarizes studies which show that people believe in hypotheses that, if true, would explain an event, but also refers to studies which show that this effect can be undone if a person is presented with a competing hypothesis. It was discovered that availability of an alternative explanation often reduces or even eliminates the perceived truthlikeness of the initial explanation. This result is compatible with the psychological hypothesis. First, it shows that a comparison among explanations plays a role in determining the perceived probability of these explanations. Secondly, in the absence of alternative explanations a mere ability to explain is sufficient for accepting of the hypothesis. The theories of IBE only require that this mere ability to explain should be good enough.

Another experiment by Read and Marcus-Newhall (1993) showed that perceived probability of an explanation depends not only on the availability of alternative explanations, but also on their perceived explanatory goodness. Read and Marcus-Newhall observed that the perceived goodness of a set of narrow explanations was lower when a broad explanation was present than when it was absent. Hence, the presence of a better explanation lowered the perceived probability of other explanations.

IBE appears to be psychologically adequate: the experimental results mentioned in this section show that people exhibit different features of IBE in their reasoning. Even though these results cannot be said to be conclusive, they do give empirical support for the truth of the psychological hypothesis which is applied by the pragmatic and evolutionary justifications of IBE.

2.3.2 Pragmaticism and IBE

There is no philosophical controversy that IBE is pragmatically warranted or at least pragmatically motivated. Peirce not only introduced the concept of abduction, but also the conception of pragmatism. He put forward the following pragmatic maxim:

> consider what effects, that might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of these effects is the whole of our conception of the object. (Peirce 1934: 5.2)

Hence, anything is conceivable or has a meaning for a pragmatist if and only if it has conceivable practical effects.

Moreover, for Peirce, pragmatism is the logic of abduction. Abductive inference is an acceptable explanation of the abductive trigger only if it explains the abductive trigger and if and only if it is susceptible to experimental verification (Peirce 1934: 5.197). The latter requirement is equivalent to the requirement of the pragmatic maxim that anything conceivable has to have conceivable practical effects. Hence, pragmatism is the logic of abduction, because something can be accepted as an abductive inference only if it satisfies the pragmatic maxim.

Abduction is justified for Peirce, because abduction is the only means to introduce a new idea, to learn or to understand new things. The required possibility of experimental verification safeguards that false abductive conclusions would not be accepted as beliefs. As Peirce himself puts it, abduction is

> the only logical operation which introduces any new idea,
>
> if we are ever to learn anything or to understand phenomena at all, it must be by abduction that this is to

> be brought about (Peirce 1934: 5.171),
>
> its only justification is that if we are ever to understand things at all, it must be in that way (Peirce 1934: 5.145),
>
> its justification being that it is the only possible hope of regulating our future conduct rationally (Peirce 1932: 2.270),

and

> its only justification is that from its suggestion deduction can draw a prediction which can be tested by induction. (Peirce 1934: 5.171)

This kind of justification for a pragmatist is epistemic justification. According to Almeder (2007), pragmatists may accept some beliefs as epistemologically justified even if they are not justified by inductive or deductive inference from already justified beliefs, but they satisfy certain other conditions. Almeder distinguishes two such conditions. More particularly, pragmatism is defined by a principle:

> A person will be rationally justified in accepting a proposed proposition P as true if
>
> (a) After exhaustive research, there is at that time no currently available conscious inference, either inductive or deductive, from other antecedently known or justified beliefs that would either confirm or disconfirm the proposition P; and
>
> (b) There is some real possibility that accepting P as true, or very likely to be true, will have a tendency to provide behavioral consequences more productive of cognitive or moral utilities than would be the case if one

had accepted instead either the denial of P or nothing at all. (Almeder 2007: 172)

If we apply these conditions in the context of IBE, we get conditions under which IBE is pragmatically epistemically justified. The condition (a) is equivalent to the requirement that we infer the best explanation among the competing explanations. The condition (b), in its turn, requires that the best explanation facilitate cognitively or morally useful behavioral consequences. So any IBE would be pragmatically epistemically justified if in addition to being the best explanation it would also be somehow useful or valuable.

There are several reasons how IBE can be said to be useful. First of all, the better the explanation, the easier it is to understand, to use or to operate with. The best explanation by definition is the one that provides the most understanding and it is easier to work with something one understands well than with something one lacks understanding of. More particularly, the best explanation is discerned with the help of explanatory virtues and each of the explanatory virtues prefers hypotheses that are easier to understand and work with. If a hypothesis coheres well with background knowledge it is easier to understand because of the relation to the already known things. Simpler and shorter hypotheses are usually easier to understand than longer and more complex ones. It is easier that one can apply the same explanatory pattern in many areas than to have separate explanations for every problem. Having a further deeper explanation for some explanatory primitives lets us use those primitives with more understanding and with greater ease. Finally, IBE eliminates competing hypotheses and thus provides us with more compact and simpler set of hypotheses to work with; however a selection of an arbitrary hypothesis would work the same way in this

respect.

Secondly, IBE is economic. To infer that something is the best explanation given the background knowledge is more economic than to search for additional knowledge that may or might not explain the abductive trigger. It is cheaper to stick to the best explanation than to allocate all the resources for the search of definite truth. This is most evident in case of abduction. According to Gabbay and Woods,

> abduction is a procedure in which something that lacks epistemic virtue is accepted because it has virtue of another kind. (Gabbay and Woods 2005: 62)

What other kind of virtue abduction possesses is explained by Floridi:

> The bet can be risky (we may be wrong), but it often pays back handsomely in terms of lower amount of informational resources needed to reach a conclusion. (Floridi 2009: 322–323)

A similar conclusion is made by Kelly (2007) who claims that choosing the simplest hypothesis compatible with experience and hanging onto it while it remains the simplest is both necessary and sufficient for efficiency, i.e., for minimizing the total number of times one gets a false inference prior to convergence to the true inference.

Finally, the truth of psychological hypothesis together with the evolutionary psychology interpretation of it implies that IBE is pragmatically warranted. According to evolutionary psychology, we can expect organisms to possess and entertain a psychological trait only if it contributes to survival and reproduction, i.e., if it is a product of the natural selection (or at least its byproduct). Similar arguments are used to justify the reliability of IBE (Carruthers 1992: ch. 7;

Goldman 1990), abduction (Floridi 2009: 324; Peirce 1932: 2.753) or beauty in explanations (Kuipers 2002: 300–301). The use of IBE among people means that IBE is a reliable ampliative method of reasoning. As Carruthers states it,

> it is hard to see in what other way inference to the best explanation could have survival-value, unless it is indeed reliable. (Carruthers 1992: 110)

Its us e helped to survive and helps to successfully act in the world. If a more reliable ampliative method of reasoning were known it would have displaced IBE and became dominant in the course of evolution. This does not imply that IBE is infallible, but only that other ampliative methods should be even more erroneous. If IBE is really a product of natural selection it means that it is reliable to the extent that there is no other known ampliative method of reasoning whose success rate would be as high as that of IBE. This implies a strong pragmatic justification. A belief is pragmatically justified if an action based on that belief would be successful. According to the evolutionary psychology interpretation, people use IBE because there is no other more reliable ampliative method of reasoning than IBE. Hence, beliefs acquired by IBE have to be successful more often than beliefs made by any other ampliative method. If one wants to act as successfully as possible the evolution suggests that she should employ IBE to form her beliefs.

Enoch and Schechter (2008) claim that pragmatic justification is an epistemic justification only for the basic belief forming methods, i.e., methods that are indispensable and rationally required to pursue. Their use and value is in their indispensability. Enoch and Schechter treat IBE as a basic belief forming method and, therefore, as a justified one. Goldman even claims that

> creatures otherwise like us would not have survived without the capacity for such knowledge and the inferential ability that it requires. (Goldman 1990: 40)

Therefore, IBE is considered to be pragmatically justified even under the more stringent conditions put forward by Enoch and Schechter.

Nonetheless, even if IBE is pragmatically warranted, it does entail that IBE is truth-conducive. For example, simpler and more unifying theories tend to make more accurate predictions if they are intentionally designed not to perfectly fit the data (Forster and Sober 1994). Some abductions are made contrary to the facts (Gabbay and Woods 2005: 119–120), i.e., abduction can be radically instrumentalist in a sense that an explanatory hypothesis may be inferred even though one knows in advance that it cannot be true. Any instrumentally successful theory is pragmatically warranted, but it is well known that instrumentalism is not a good friend with scientific realism. Therefore we should agree with Schurz who claims:

> I regard the instrumentalistic perspective as an important warning that not every empirically useful theoretical superstructure must correspond to an existing structure of reality. (Schurz 2008: 231)

This means that the pragmatic warrant cannot and does not intend to grant the truth-conduciveness of IBE. When a pragmatists claims that IBE is epistemically justified, the term "epistemically" in the pragmatist's claim has nothing to do with truth or truth-conduciveness.

2.3.3 Evolutionary Justification

Even though pragmatic justification does not entail that IBE is truth-conducive, some philosophers elaborate pragmatic justifica-

tion and claim that the evolutionary psychology interpretation of IBE as an actual cognitive capacity entails that the conclusions of IBE are true. Quine is the first who put forward an argument of this kind:

> creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind. (Quine 1969: 126)

A much more sophisticated evolutionary argument is presented by Boulter (2007). According to him, organisms that can successfully interact with their environment have a higher chance of reaching maturity and reproducing their own kind. Beliefs about their environment that accurately track the state of affairs are, on the whole, better guides to action than are false beliefs. Hence, natural selection will favor organisms with reliable sensory and belief formation systems insofar as those systems have a direct bearing on the organism's ecological and social fitness. If false and true beliefs have a direct bearing on human ecological and social fitness then it is not likely that people will tend to believe what is obviously false or fail to believe what is obviously true. Common-sense beliefs have a direct bearing on human ecological and social fitness; therefore common-sense beliefs will tend to be correct. Nevertheless, these two arguments do not directly apply to IBE. Quine's argument applies to induction in general and thus it applies to IBE only indirectly. Boulter's argument applies only to common-sense beliefs and only to those common-sense beliefs that have a direct bearing on the animal's ecological and social fitness.

Carruthers (1992) and Goldman (1990) apply the evolutionary argument directly to IBE by arguing that IBE provides true beliefs. Carruthers is rather straightforward. He claims that IBE is an innate

human cognitive capacity, because people posses it even though they are not explicitly taught to do it and because it does not seem to be learned from experience (Carruthers 1992: ch. 7). He further claims that the reliability of IBE easily explains why it is the case that IBE is innate. IBE is innate, because it is reliable, i.e., it provides true beliefs, and because of this reliability the capacity for IBE prevailed by means of natural selection:

> individuals will be better able to survive if they are able to attain true beliefs about the underlying processes at work in nature, which can then be harnessed and exploited, or if they can acquire knowledge of the unseen causes of observable phenomena. (Carruthers 1992: 110)

Carruthers (1992: 184) thinks it is very unlikely that the survival value of explanatory power can be explained away in terms unrelated to truth.

Goldman's argument is more cautious and elaborate. A cognitively relevant unit of natural selection can only be a set of genes or alleles responsible for brains of a certain size and structure. Our brains and, consequently, the genes that produce it, may have been selected because of certain capacities, including the capacity for a certain inference pattern, namely IBE. However, the existence and usefulness of some capacity does not entail that it was selected. Goldman maintains that it is not the case for IBE, because IBE is used universally and nearly uniformly throughout our species, because rudiments of IBE are found in biologically related species that evolved earlier and because of the high chance that IBE is necessary for the survival of a species. Secondly, the natural selection of IBE might not entail that it produces true beliefs. According to Goldman this is not the case because an ability to infer to the actual

sources of food, danger and other perceptual knowledge on the basis of perceptual data clearly enhances fitness. Therefore IBE preserves truth at least in these contexts. Finally, the production of true beliefs in the latter contexts does not entail that it will do so in other contexts. More particularly, it is not clear if IBE preserves truth in contexts beyond the perceptual and observable, i.e., in the context of unobservable. Goldman maintains that IBE preserves the truth in these contexts too. Firstly, because there may be no sharp and fixed epistemic distinction between observable and unobservable, or between cases relevant in natural selection (danger or food) and cases that are not relevant. Secondly, experimental testing of scientific theories uses different apparatuses and procedures that provide results in accordance with particular theories about the unobservable. Therefore, experimental testing is efficacious enough to select from among competing explanatory alternatives.

Quine, Boulter, Carruthers and Goldman claim that evolutionary psychology entails the truth-conduciveness of our cognitive abilities. However, other philosophers argue that these claims are underdetermined. The most common objection states that the ultimate purpose and function of cognitive capacities that originated through natural selection is survival and it is not evident whether it includes the generation of true beliefs. Boulter, Carruthers and Goldman do not deny that. They rather claim that the generation and possession of true beliefs makes survival much more easily attainable and explainable than the generation and possession of false beliefs. Therefore, the argument states that even though the ultimate goal of natural selection is survival, true beliefs rather than false ones facilitate the survival. This argument for the truthlikeness of evolutionarily achieved cognitive capacities will be blocked if there are survival-facilitating beliefs that are actually false.

The most evident example of these kinds of beliefs is the so-called "better-safe-than-sorry" beliefs:

> a very cautious, risk-aversive inferential strategy—one that leaps to the conclusion that danger is present on very slight evidence—will typically lead to false beliefs more often, and true ones less often, than a less hair-trigger one that waits for more evidence before rendering a judgment. Nonetheless, the unreliable, error-prone, risk-aversive strategy may well be favored by natural selection. For natural selection does not care about truth; it cares only about reproductive success. And from the point of view of reproductive success, it is often better to be safe (and wrong) than sorry (Stich 1990: 62).

This will occur more often the cheaper the false positive beliefs are and the deadlier the false negative beliefs are. More particularly, the rate of false beliefs should be inversely proportional to the cost of the false positive beliefs and directly proportional to the cost of the false negative beliefs. Pascal's wager is also an instance of the better-safe-than-sorry belief. It is safer to bet for God and gain an infinitely happy life (if God exists) or lose almost nothing (if God does not exist) than to bet against God and gain very little (if God does not exist) or gain nothing (if God exists). In a similar vein, it is safer for an organism to flee at the slightest sign of danger (whether it is really the case or not) than to die because of the underestimation of the danger. Similarly, there is a good evolutionary explanation for why Thelma believes that her children are more beautiful and smarter than average. However, this belief is not properly causally connected to the objective qualities of her children. Therefore, Thelma's belief is survival enhancing for her

children, but can be objectively false (De Smedt and De Cruz 2010). Stephens (2001) shows that the better-safe-than-sorry beliefs have a very narrow application, nevertheless his own model also indicates that natural selection will not always favor true beliefs.

Boulter (2007: 377) notices that too many better-safe-than-sorry beliefs will be positively maladaptive since they would prevent one from engaging in important activities. Nevertheless, this counter argument cannot deny the possibility of false but survival enhancing beliefs. Moreover, if survival does not require true beliefs, other contexts might not require them either. Therefore, the evolutionary justification cannot grant that IBE is truth-conducive.

2.4 Probabilistic Justification

There are endeavors to explicate particular explanatory virtues as truth-conducive or, at least, conformation-conducive. One can analyze probabilistically the truth-conduciveness of coherence, unification and simplicity.

Coherence. Most generally, IBE is an inference to the hypothesis that is most coherent with background knowledge (1.3.1). However, it is probabilistically impossible that more coherent set of propositions were more probable than less coherent one. Klein and Warfield (1994) presented two premises that jointly entail the impossibility result:

> a more coherent set of beliefs resulting from the addition of a belief to a less coherent set of beliefs is less likely to be true than the less coherent set of beliefs. (Klein and Warfield 1994: 130)

The first premise states that a consistent set of beliefs $B = \{p, q\}$ is more probable than any set B^*, which contains all members of B

and one additional proposition r, i.e., $B^* = \{p, q, r\}$, so long as r has neither an objective probability of 1, i.e., $Pr(r) < 1$, nor is entailed by B, i.e., $B \nvDash r$. In other words, according to probability theory B^* cannot be more probable than B, i.e., $Pr(B^*) \leq Pr(B)$. The second premise states that one strategy of converting a less coherent set of beliefs B into a more coherent set of beliefs B^* is to add a belief to B that has neither an objective probability of 1 nor is entailed by B. However, as was shown in the first premise, according to probability theory B^* cannot be more probable than B. Therefore, a more coherent set of beliefs cannot be more probable.

Bovens and Olsson (2002) argue that this negative conclusion is false. Firstly, Klein and Warfield are claimed to improperly distinguish between the sets of beliefs and the sets of propositions. Their result holds for the sets of propositions, but not necessarily for the sets of beliefs, which are also called information sets, testimonial systems or belief systems. A belief system differs from a simple set of propositions in a way that a belief system is a set of propositions believed by some particular person. Therefore, whereas the probability of a set of propositions is equal to the product of their probability, the probability of a belief system is the probability that these propositions are all true given that they all are believed by the relevant person reported by a witness or gathered by some other means. There are also additional constraints ascribed to belief systems. Beliefs are defined to be gathered from partially reliable (relatively unreliable) and independent sources. Secondly, if a belief system is defined the latter way, then the crucial premise in the Klein and Warfield's argument is false. More particularly, Bovens and Olsson showed that a larger belief system need not be less probable but can be even more probable than its parts.

Thus defined is the problem of the truth conduciveness of coher-

ence. Does greater coherence among propositions in an information set, where each proposition was reported by an independent and partially reliable witness, imply greater probability of the set, ceteris paribus (i.e., when factors that have nothing to do with coherence are fixed)? There are two famous impossibility results that show that coherence is not truth conducive in this defined sense. (Bovens and Hartmann 2003: 20–21) provide an example in which every measure of coherence would rank one information set as more probable than another for some values of witness reliability and as less probable for other values. From this example they conclude that there cannot be a probabilistic measure of coherence that imposes a coherence ordering on information sets, which is fully determined by the probabilistic features of the information sets, and that would rank a more coherent set as a more probable one, ceteris paribus. Olsson (2005b;a) also shows that posterior probability of an information set (a testimonial system) depends not only on the probability of what a proposition says, but also on the probability that the proposition is reliable (or that report of this proposition is reliable). He concludes that there cannot be a non-trivial (informative) coherence measure that is truth conducive ceteris paribus in a basic Lewis scenario. Olsson uses the concept of a basic Lewis scenario to refer to a situation when relatively unreliable witnesses tell the same story and where this coherence, according to Lewis (1946: 246), indicates the high probability that of what the witnesses agree upon.

The impossibility results claims that coherence is not truth conducive ceteris paribus. There are factors that are supposed to have nothing to do with coherence and therefore are held fixed. Firstly, both Bovens and Hartmann (2003: 11-12) and Olsson (2005b: 395, 404) hold that the reliability of information sources among information sets should be fixed as equal. This is required, because,

intuitively, information set acquired from a more reliable reports would seem to be more probable than one acquired from less reliable sources. Secondly, Bovens and Hartmann (2003: 11-12) require that the expectedness or a perceived prior probability of information sets should be fixed equaly, because one wants to measure the effect of coherence on the confidence in the information sets and different expectedness can bias the confidence. Olsson (2005b: 395, 404), on the other hand, does not think that expectedness or prior probability should be fixed equal, because coherence can depend on the prior probability and some probabilistic measures of coherence are very dependent on it.

Meijs and Douven (2007: 352–353) propose that maybe the truth conduciveness of coherence can be salvaged if one adopts additional or even completely different ceteris paribus conditions. Meijs and Douven themselves propose that we should accept equal witness reliability and equal marginal probabilities of the propositions in the sets as the ceteris paribus conditions and that there are no counterexamples to the truth conduciveness of coherence given these revised ceteris paribus conditions. Glass (2007: 285) notes that a measure proposed by Olsson (2002) and Glass (2002) is truth-conducive for information pairs. Hence, if one builds a set of beliefs incrementally by adding a new belief to the conjunction of the previously evaluated beliefs, then coherence of a set with only two elements is sufficient to analyze the truth conduciveness of coherence. Bovens and Hartmann (2003) themselves propose a quasi or partial ordering of coherence on the set of information sets. According to this measure, for a set of information sets with the same cardinality and equal prior joint probability, the posterior joint probability of an information set S is no less probable than of S' if and only if S is no less coherent than S' for all values of the reliability parameter. Hence, maybe there are

ways in which coherence as defined by the probabilistic measures of coherence is truth conducive. However, it seems that neither the impossibility results, nor ways to escape these results are actually applicable to the theory of IBE.

First of all, the very results Klein and Warfield (1994) obtained are not applicable to the theory of IBE. Klein and Warfield argue that a larger set of belief cannot be more probable than a smaller one even if it is a more coherent one. The theories of IBE, on the other hand, are interested in the truth of only one element of the coherent set, i.e., an explanatory hypothesis, and not in the truth of the whole set. Merricks (1995) points out that the result obtained by Klein and Warfield does not imply that coherence cannot be truth conducive for a particular belief from a coherent set rather than for the whole coherent sets of beliefs.

Furthermore, the impossibility results obtained by Bovens and Hartmann and Olsson are not applicable to the theory of IBE either. This is the case, because the required conditions do not hold in the case of IBE. There is no independence. A hypothesis is better the more it is connected with other beliefs in background knowledge, the more probable it is given the background knowledge, and the more probable the background knowledge is given the hypothesis. This cannot be the case if the hypothesis and background knowledge were independent. Moreover, in the MMH measure of unification or in a case of application of CCP, an explanatory hypothesis unifies some data exactly if they, respectively, make the data dependent or independent.

There is no partial reliability. The set to be evaluated consists of an explanatory hypothesis, background knowledge and an abductive trigger, which is an element of background knowledge. Background knowledge is taken to be true; hence it is taken to be totally reliable.

The explanatory hypothesis, on the one hand, is partially reliable, because it is the product of abduction (a partially (sometimes) reliable process of inference), on the other hand, it cannot be ascribed reliability or unreliability, because it is not reported, but generated. One does not hold it as a belief, but only tentatively accepts it in order to conduct a comparison of explanatory power.

There is no equal prior probability. Neither the joint prior probability of a set, nor the marginal prior probability of the elements of the set can be held equal in the case of IBE. The majority of the elements of competing sets, i.e., background knowledge or abductive triggers, are the same and are held to be true, i.e., their probability is considered to be equal or close to 1. Competing sets differ only in explanatory hypotheses. Hence, the joint probability of the sets depends entirely on the marginal prior probability of the explanatory hypotheses. But the marginal prior probability of explanatory hypotheses cannot be held equal in an evaluation of coherence, because the former is claimed to be a function of the latter. If the theories of IBE are correct, and then if we hold prior probability (joint or marginal) equal, we would not be able to differentiate coherence.

Therefore the results of the analysis of the truth-conduciveness of coherence in epistemology are not applicable to IBE, because different assumptions that are made in this analysis are incompatible with the theories of IBE.

Glass (2010) tested a different approach to the problem at hand. He made a computer simulation of different probabilistic measures of explanatory power in order to find how well the measures will identify the explanation which has been designated to be the actual hypothesis. Glass compared measures of explanatory power based on MPOST, ML, the conservative Bayesian approach, the difference measure of confirmation $d(H, E) = Pr(H|E) - Pr(H)$, the likelihood

ratio measure of confirmation $l(H, E) = log(Pr(E|H)/Pr(E|\neg H))$, the overlap measure of coherence $C(H, E)$ (E_G) and the Fitelson (2003) measure of coherence.

Two sets of simulations were conducted: fair and biased. The instruction for the fair simulations is the following:

1. Randomly assign prior probabilities to each hypothesis H_i. These probabilities are constrained to sum to one.
2. Randomly assign a likelihood $Pr(E|H_i)$ to each hypothesis.
3. Randomly select one of the hypotheses using the prior probability distribution and designate this hypothesis as the actual hypothesis H_A.
4. Select whether E or $\neg E$ occurs using the likelihood of H_A so that there is a probability $Pr(E|H_A)$ of E occurring and $1 - Pr(E|H_A)$ of $\neg E$ occurring.
5. For each approach, if E occurs, identify which hypothesis provides the best explanation of E, and similarly if $\neg E$ occurs.
6. For each approach, if the hypothesis identified in step 5 matches the actual hypothesis, count this as a success, otherwise count it as a fail.
7. Repeat steps 1 to 6 to obtain an accurate value for the percentage of successes (accuracy) for each approach.

The fair simulations were conducted with sets of 2-10 mutually exclusive and exhaustive competing explanations. The biased simulation differs from the fair one at step 5:

5. (a) Introduce a random error to the prior probabilities assigned in step 1 by adding a number sampled

from a Gaussian distribution with mean zero and a specified standard deviation, provided the resulting probability lies in the interval [0, 1]. If it does not, the process should be repeated until it does.

(b) Normalise the probabilities resulting from the previous step to ensure they sum to one.

(c) For each approach, if E occurs, identify which hypothesis provides the best explanation of E, and similarly if $\neg E$ occurs.

The biased simulations were conducted with sets of 2 mutually exclusive and exhaustive competing explanations. The simulations were different among themselves in the values of the applied standard deviation. In the biased simulation only the measures MPOST, ML and E_G were evaluated, because ML is equivalent with the rest of the measures when only two hypotheses are being compared.

Glass observed that MPOST perform best out of the evaluated approaches in the fair scenario, however he rejects MPOST as an account of best explanation because it simply choses the most probable as the best explanation without any genuine account of explanatory power. In other words, MPOST as an account of explanatory power would make IBE trivial (1.4.2). Glass takes MPOST to be a benchmark (because it defines what the most probable explanation is) against which all the other measures have to be compared. As the uncertainty (the standard deviation) is introduced and increases in the biased scenario, the results for PMOST decline. MPOST remains the best at the values of 0–0.4 of a standard deviation. E_G is the best at the values of 0.4–0.7 and for a standard deviation of 0.7 and higher ML outperforms both MPOST and E_G. Given these results Glass concludes that E_G is the most plausible account

of the best explanation among the evaluated measures. In the fair scenario it identified the best explanation almost as often as the MPOST measure (which is rejected as an account of the best explanation by Glass because it would trivialize the concept of the best explanation). In the biased scenario, the average of its success was higher than of the rest of the measures.

If IBE is truth-conducive, then the best explanation will be the most probable explanation. MPOST defines what it takes to be the most probable explanation or hypothesis. The measure E_G identifies the actual explanation almost as often (98.6–98.9% as often) as MPOST. Hence, if the measure E_G is an adequate measure of explanatory power, then the simulations show that IBE is truth-conducive. More particularly, firstly, to the extent that the measure E_G is an adequate account of coherence and to the extent that it succeeded better than other possible probabilistic accounts of explanatory power, the result vindicates the idea that if one wants IBE to be truth-conducive, then IBE should be defined as the inference to the most coherent explanation. Secondly, to the extent that the measure E_G as a measure of coherence is an adequate account of explanatory power, the results suggest that IBE is truth-conducive.

However, this result should not comfort the proponents of IBE. The result cannot be considered as proof of the truth-conduciveness of IBE. The point has to do with the insufficiency of probabilistic considerations in the context of IBE. Probabilistic measures of explanatory power work best—they correctly identify the actual explanation—if the correct probability distributions are fed into them. For example, the simulations reveal that MPOST is the best if we have the correct distribution and the E_G measure is on average better if we have biased prior probability distributions (unfortunately, Glass did not test what would happen if the likelihoods

were also biased). Explanationists claim that explanatory considerations can or ought to facilitate the determination of probability distribution. The problem of the truth-conduciveness of IBE then is whether explanatory considerations provide the correct probability distribution. More particularly, whether the most powerful explanation is the one that has the highest prior probability and likelihood. If IBE is truth-conducive, then the application of IBE would provide the correct probability distribution. However, if we have the correct probability distribution, there is no need for the measure E_G, because MPOST would be the best means to identify the most probable explanation. The real problem of IBE thus is not to find the best explanation after we somehow got the correct probabilities, but whether we can have the correct probabilities in the first place and the measure E_G is of no use here. In other words, the measure E_G (as any other probabilistic measure of explanatory power) is only of use after all the work of evaluation of explanatory power has been done. All this once again stresses that probabilistic accounts of explanatory power are insufficient and that the explanationist approach is more fundamental than the probabilistic one.

Explanationists want to employ explanatory virtues to determine the probability distribution. Therefore, we can next examine if particular virtues that constitute coherence are truth-conducive.

Unification. Unification is neither sufficient nor necessary for explanation. However, it is one of the most important and often-cited explanatory virtues. The question then is whether a unifying hypothesis will also be more probable than less unifying one. If we measure unification with the MMH measure the answer to this question is positive. A more unifying hypothesis H_1 will have higher posterior probability than the less unifying hypothesis H_2:

$$Pr(H_1|E_1 \wedge \ldots \wedge E_n) > Pr(H_2|E_1 \wedge \ldots \wedge E_n).$$

In the words of McGrew,

> if we change our focus from the independence of the evidence apart from the theory to its dependence in light of the theory, we discover a lovely theorem: the degree of confirmation a hypothesis receives from the conjunction of independent pieces of evidence is a monotonic function of the extent to which those pieces of evidence can be seen to be positively relevant to each other in the light of that hypothesis. (McGrew 2003: 561–562)

However, we have argued that the MMH measure of unification is not adequate, at least in respect to IBE. Unification is not a matter of degree: it is a matter of how many different kinds of data a hypothesis unifies.

A more unifying hypothesis can get evidential support from a bigger number of phenomena. Suppose the hypothesis H_1 explains only E_1 and, thus, gets evidential support only from that. Another hypothesis H_2 explains only E_2 and also gets evidential support only from this piece of evidence. A unifying hypothesis H_u that explains two different kinds of data d_1 and E_2, on the other hand, gets evidential support from both E_1 and E_2 at the same time. However, a conjunction of H_1 and H_2 will be supported by both E_1 and E_2, hence it would be equally supported as H_u.

What is more important, the content of a not-unifying hypothesis H_1 or H_2 will always be more probable than the content of a unifying hypothesis H_u. More particularly, H_1 claims that only E_1 is the case and H_u claims that both E_1 and E_2 is the case. By the probability theory, $Pr(E_1) \geq Pr(E_1 \wedge E_2)$. Hence, what H_u claims is less probable than what H_1 claims and, consequently, the truth of H_u is less likely than the truth of H_1. This is exactly the claim

Popper (2002/1959) made that a theory with greater empirical content cannot be more probable than a theory with a smaller empirical content.

Explanatory power should provide understanding. Humphreys (1993) shows that this is not necessarily the case. Consider two different axiomatizations of the propositional logic L

$$((((A \to B) \to (\neg C \to \neg D)) \to C) \to E) \to$$

$$\to ((E \to A) \to (D \to A))$$

and L'

$$A \to (B \to A),$$

$$(A \to (B \to C)) \to ((A \to B) \to (A \to C)),$$

$$(\neg B \to \neg A) \to ((\neg B \to A) \to B).$$

L and L' language is the same, they have the same connectives, well-formed formulas are defined the same way, each has the same rule of inference (modus ponens) and, because both are complete axiomatizations of propositional logic, the sets of their logical consequences are the same. L is more unifying, because it consist of only one axiom rather than three as L' does, and hence should provide greater understanding. However, L is less intuitive and it is harder to understand than L'.

Unification is one of the most important explanatory virtues. However, it is also one of the most metaphysical. There is no a priori reason to suppose that the world is unified rather than disunified and complex. Probabilistic considerations are indifferent in relation to this question or even slightly against the truth conduciveness of unification. Unification will be truth conducive only if the world is unified and this can only be established empirically, if at all.

Simplicity. Probabilistically, a hypothesis consisting of fewer propositions is more probable than a hypothesis consisting of more propositions. That is a hypothesis consisting of propositions A and B is more probable than a hypothesis consisting of A, B and C. This might not be the case if the hypothesis consisting of fewer propositions is not a proper subset of the hypothesis consisting of more propositions. But for the latter to hold, one has to have prior probability distributions for the hypotheses at hand. If one does not have prior probability distributions ascribed to the hypotheses, then, by the principle of indifference, the hypotheses consisting of fewer propositions have to be considered more probable. Moreover, complex hypotheses or hypotheses with more adjustable parameters can be true in more ways than simpler hypotheses or hypotheses with fewer adjustable parameters. Hence, a hypothesis with more different ways to be true, by the principle of indifference, is less probable in every instance of the way it can be true than a hypothesis that can be true in fewer different ways. However, as the Bertrand paradox shows, the principle of indifference can facilitate the ascription of different contradictory prior probability distributions.

A more parsimonious hypothesis can be more difficult to understand than a more complex one. Analogously, as in the case of unification, simplicity does not guarantee easier understanding. Humphreys' (1993: 184) example with two different axiomatizations of propositional logic shows that paucity of assumptions does not entail easier understanding. Similarly, Barnes notes that

> complex and inelegant theories often offer rich understanding, where causal histories of explananda happen to be messy and complicated (consider explananda like 'Joe eventually married Sally' or 'World War II ended in 1945'). (Barnes 1995: 265–266)

This is tightly connected with claims that there are contexts where simplicity is less probable than complexity. For example, social sciences (Salmon 2001b: 129) or biology (Carruthers 2006: 151) are claimed to be such contexts.

Simplicity is not only the most difficult of the explanatory virtues to define, but also the most difficult one to asses. Simplicity is the most likely one of the explanatory virtues to produce incommensurable orderings of explanatory power. It is not always possible to adjudicate the simplicity of two different hypotheses. Competing hypotheses might not be content-comparable (Grünbaum 2007). Moreover, whether something is simple depends on how it was defined. One can always use the most conveniently simple domain and vocabulary (Kuipers 2002: 302) or, as Lakatos claims,

> no doubt, simplicity can always be defined for any pair of theories T_1 and T_2 in such a way that the simplicity of T_1 is greater than that of T_2. (Lakatos 1970: 131 note 106)

Because of all this relativity it can be very difficult to prove that one rather than another ordering of simplicity is the correct one.

Simpler hypotheses (ones that minimize the number of adjustable parameters) make better predictions in the curve-fitting problem. However, simpler hypotheses in the curve-fitting problem are known not to be totally accurate. This way, on the one hand, they do not perfectly fit the data, but, on the other hand, they do not overfit the noise in the data, and because of that they make more accurate predictions. Hence simplicity is more a means to escape noise in data than an indicator of truth of the hypothesis. In the curve-fitting problem simplicity is prediction-conducive and not very truth-conducive. Moreover, simplicity in curve-fitting is significant only if the data set is small. If the data set becomes bigger, the need for

simplicity ceases.

A related point is that even if it is not clear whether simplicity is truth-conducive, it is at least efficiency-conducive. Kelly (2007) showed that sticking to the simplest hypothesis is the most efficient way to the truth:

> the best that Ockham's razor could guarantee a priori is to keep us on the straightest possible path to the truth, allowing for unavoidable twists and turns along the way as new effects are discovered—and that is just what it does guarantee. (Kelly 2007: 563)

Analogously with unification, simplicity is one of the most cited explanatory virtues, but also one (together with unification) of the most metaphysical. There is no a priori reason to suppose that the world is simple rather than complex. Hence, semantic simplicity earns the same judgement as unification. Moreover, against the truth conduciveness of simplicity plays the fact that simplicity is very relative to the vocabulary in which the relevant hypothesis is expressed. Hence, syntactic simplicity is even less likely to be truth conducive than semantic one. All in all, the above conceptual arguments judge more against the truth-conduciveness of simplicity.

The Impossibility of Probabilistic Justification. Probabilistic measures of explanatory power are functions of likelihoods and prior probabilities or can be expressed this way. When expressed this way it follows trivially that the best explanation among the competing hypotheses would also be the most probable one, i.e., the hypothesis with the highest posterior probability. This would happen because both explanatory power and posterior probability would be the direct functions of the same arguments (the probability of the evidence is also an argument of posterior probability or can

appear in some of the probabilistic measures of explanatory power, however when evaluating competing explanations for the same piece of data the probability of the evidence remains constant and therefore does not have any impact on the value calculated by the measures). Hence, if we accept any of the so far discussed probabilistic measures of explanatory power (1.4.1 and 2.4), loveliness will lead to likeliness. However, it will happen in a rather trivial way.

This result is undesirable not only because of the triviality, but also because the probabilistic measures of explanatory power utilize likelihoods and prior probabilities when these should be better seen as determined by the explanatory considerations (1.4.3). Moreover, probabilistic measures of explanatory power seem to be insensitive to differences in explanatory virtues. That is, competing explanatory hypotheses can have the same likelihoods, prior probabilities or even the final values of explanatory power, but differ in the degree of unification, depth or simplicity they provide. In this kind of situation these hypotheses would intuitively differ in explanatory power, but probabilistically would have the same value of explanatory power.

Psillos denies that IBE can be put in a neat formal form and justified deductively:

> in an ampliative method the alleged transference of the epistemic warrant from the premises to the conclusion depends on substantive (and hence challengeable) background beliefs and considerations. (Psillos 2002: 608)

Douven and Horsten explain why this is the case:

> considering all possible models means testing in a knowledge vacuum. (Douven and Horsten 1998: 316)

Therefore, the pivotal role of background knowledge in IBE is the main obstacle to the formal justification of IBE. The function of IBE

is creative (ampliative). Ampliative inferences are non-deductive. The formal justification of IBE would show that IBE is deductive. Deductive inferences are non-creative (non-ampliative). Therefore, if the theories of IBE do not want to become contradictory, they have to reject the possibility of a formal or probabilistic (which is an instance of a formal justification) justification of IBE.

2.5 Ontological Commitments and Falsification

2.5.1 Ontological Commitments of IBE

This section enumerates ontological commitments that the theories of IBE make. These commitments can then be tested historically for whether theories accepted as true are the ones that conform to the criteria of a good explanation, i.e., whether these theories unfold the relevant ontological order.

IBE is not deductively valid and, therefore, best explanations cannot be true in every possible world. According to Lipton:

> unlike the principles of deductive inference, reliable principles of induction are contingent. [...] A pattern of non-demonstrative inference that generally takes us from truth to truth in this world would not do so in some other possible worlds. (Lipton 1993: 101)

Nevertheless, IBE is argued to be truth-conducive: reliable and providing true conclusions. IBE is a material form of inference, not formal. Its validity depends not on the form of the inference, but on its content and on the meanings of the terms it employs. In other words, there are substantive assumptions that have to hold for IBE to be truth-conducive and that have to hold if IBE is truth-conducive. These assumptions have to do with the way our tangible

world is. Some of them are characteristic of induction in general and some are characteristic solely of IBE, but all of them are the claims about the ontological structure of the world. Even if it were possible to establish the truth-conduciveness of IBE formally it would still have ontological consequences. A formal proof of the truth-conduciveness of IBE would mean that IBE guarantees truth in all possible worlds and, obviously, in specific one among them, i.e., our actual world.

The biconditional explicates the connection between IBE and the ontology of the world:

> IBE is truth-conducive if and only if the actual world has a particular (coherent, unified and simple) ontological structure.

This biconditional means that the theories of IBE cannot be true and at the same time be independent of any ontological commitments. The 'if' direction clearly holds. It states a sufficient condition that would make IBE truth-conducive. It could be false, for example, only if the world was as coherent, unified and simple as possible, but explanations that are the most coherent, unified and simple would be false in that world. This seems hardly possible. The 'only if' direction states the necessary condition for IBE to be truth-conducive. After contraposition it states that "If IBE is truth-conducive the world has a particular ontological structure". For example, it can be false only if all the most coherent, unified and simple explanations were actually true, but the world would not be coherent, unified and simple. This also seems hardly possible. Moreover, the latter direction seems to be considered more characteristic of IBE than the former. For Psillos, a defeasible and ampliative type of inference, of which IBE is an instance,

> works, (it produces truths or likely truths), only if the external circumstances are right (if the world co-operates). (Psillos 2007: 442)

or

> what matters for the correctness of the conclusion is whether or not the rule is reliable that is, whether or not the contingent assumptions which are required to be in place in order for the rule to be reliable are in fact in place. (Psillos 1999: 83)

Day and Kincaid also refer to substantive assumptions as necessary for IBE to succeed:

> without substantive assumptions both about explanation in general and about specific empirical details, IBE is empty. In short, appeals to the best explanation are really implicit appeals to substantive empirical assumptions, not to some privileged form of inference. It is the substantive assumptions that do the real work. (Day and Kincaid 1994: 282)

Thus, substantive assumptions as described in these quotes are seen as necessary, without which IBE would not work. These assumptions, on the one hand, are prerequisites for IBE to work and, on the other hand, are consequences that have to follow if IBE is truth-conducive. As it is seen from the quote, Day and Kincaid even describe IBE as nothing more than the totality of these assumptions taken together.

Thagard (2007b: 29–32) gives an argument against the coherence theory of truth that is applicable here to argue for the connection between IBE and the structure of the world. Thagard claims that

historical evidence suggests the world is independent of the representation of it, and because of that the aim of representations should be the correct description of the world, not just a coherent relation to other representations. Respectively, it would be very lovely if all the true explanations were coherent and very simple, but this would not be true if the world is constituted the opposite way. IBE as formulated to date would not work in every possible world. The world has to have a very specific ontological structure for IBE to be truth-conducive. Hence, the theory of IBE is not only an epistemic and psychological theory, but also presupposes an ontological one.

Realism about the external world is the fundamental assumption of IBE. If there were no external tangible world then there would be no possibility for abductive triggers to occur, there would be no facts or events to explain. Respectively, if there were no external world then one could not state any causes (which are the favorite explanans of the proponents of IBE) that would account for the abductive triggers. All other assumptions of IBE are dependent on the realism about the external world, because all of them state how the actual world should be constituted for IBE to work.

Then there are assumptions characteristic of induction in general. These state that there are regularities in nature and that nature is uniform, i.e., the physical possibilities and regularities in nature should not change throughout space and time. Without these assumptions there could be no laws of nature and the same causes would not produce the same effects. What is more relevant for IBE, if the nature were irregular and indeterminate then any kind of explanans would be impossible, because the same explanans—even in the exactly the same circumstances—would not be capable of accouning for the same explanandum.

There are two principal substantive assumptions characteristic

solely to IBE. The first is the reliance on substantive background knowledge. In every particular instance of IBE the content of the relevant background knowledge and the truth of this content are taken as assumptions. This is one more reason why IBE cannot be truth-conducive in every possible world, because a particular content of background knowledge cannot be true in every possible world. The empirical and theoretical facts embedded in background knowledge act as the assumptions, firstly, by restricting the set of relevant and plausible hypotheses to be evaluated and, secondly, by determining the relevant explanatory considerations to be used in the evaluation. Background knowledge filters and rejects any explanation or explanatory consideration that would be incoherent or contradict it. Background knowledge also has to indicate why in a particular explanation a particular explanatory virtue contributes to the plausibility of the explanation. Moreover, acceptance of something as background knowledge implies that it is assumed to be true. Thus the best explanation can be true if and only if the particular content of background knowledge is true, i.e., if the state of affairs is exactly as described in the background knowledge.

Truth-conduciveness or confirmation-conduciveness of explanatory virtues is the second substantive assumption characteristic only to IBE. If it is really the case that each best explanation, i.e., an explanation that is more coherent, unified and simple than its competitors in a particular situation, is true, then coherence, unification and simplicity have to be truth-conducive. If IBE is a form of inference that is reliable in the actual world then the world has to be such that coherence, unification and simplicity are truth-conducive in it, i.e., it has to be coherent, unified and simple.

It was argued earlier that IBE is psychologically adequate, because people seem to follow the rule of IBE in their ampliative

reasoning. The evolutionary psychology interpretation of this fact would claim that this adaptation is due to IBE's reliability: the use of IBE helped us to survive in the world. If there were even better survival-enhancing ampliative ways of reasoning they would have replaced IBE. But, if the use of IBE helps us to successfully adapt and act in the world and there is no better kind of inference, it can indicate that IBE reflects the structure of the actual world. The theories of IBE claim that explanations that satisfy the explanatory virtues the most, i.e., are the most coherent, unified, and simple, should be accepted as true. If people are most successful when employing the most coherent, unified and simple theories that would mean that the world these people are acting in is indeed coherent, unified and simple. If the evolutionarily psychology interpretation of the psychological adequacy of IBE is true, the particular ontological structure of the actual world can be the only possible explanation for entrenchment of IBE among people as the cognitive mechanism for the ampliative reasoning (however, the better-safe–than-sorry argument refutes this interpretation).

We can now specify the biconditional put forward at the beginning of the section:

> IBE is truth-conducive if and only if there is an external world that is uniform and has regularities in it, the background knowledge depicting the state of affairs in this external world is true, and the explanatory virtues are truth-conducive.

The 'if' direction in the biconditional, being the sufficient condition, states the prerequisites that have to be true for IBE to be truth-conducive. The 'only if' direction, being the necessary condition, states the consequences that have to be true if IBE is truth-con-

ducive. The substantive assumptions are the prerequisites and the consequences of IBE at the same time. There arises a vicious circle: the only reason to believe in the truth of the assumptions that would make IBE truth-conducive is IBE itself. For example, in the realism-skepticism (about the external world) debate the hypothesis that our sense experiences are caused by the external world roughly similar to our experiences of it is taken to constitute the best explanations for these experiences (e.g., Beebe 2009). If there were no regularities, order or determinate causal-nomological structure of the world then the results and success of natural sciences would be hard to explain. In the scientific realism-antirealism debate the only reason to believe in the truth of scientific theories is the no-miracle argument (the truth of the scientific theories is the only explanation for their empirical and theoretical success) (Putnam 1975: 73), which is an instance of IBE. The background knowledge is the product of the explanatory considerations and is used at the same time to evaluate the further explanatory considerations. The particular ontological structure of the world can be the only explanation of why IBE is psychologically adequate. The proponents of IBE do not see this circle as vicious (e.g., Psillos 1999: ch. 4; Carruthers 1992: ch. 12), but rather as similar to the hermeneutical circle: IBE and its presuppositions and implications gain increasing mutual support while moving in this circle. They claim this circle is what one would expect given that the major part of justification of IBE is brought by the considerations of coherence. We are not going to evaluate the viciousness of this circle here. What is important for the task of this section is to conclude that these substantive assumptions must hold if IBE is to be truth-conducive.

Hence the theories of IBE, if true, make ontological commitments. Even the conceptual or formal establishment of truth-conduciveness

of the features of IBE will have these ontological implications, and these would have to hold in every possible world. None of the other substantive assumptions can ever be ascertained due to underdetermination. Only the empirical refutation of these claims can be conclusive. Therefore, if IBE has any non-formal ontological assumptions, we cannot ever ascertain whether those assumptions really hold in our world. The only thing we may succeed in is to ascertain, with the help of the natural sciences, that these assumptions do not hold—and this is exactly what is going to be done next.

2.5.2 Empirical-Historical Justification

The biggest approximation of a hypothesis to the truth is its empirical success. Empirical success of a hypothesis is more accessible to us than knowledge of its truth, which is, actually, inaccessible. Hence, as an indicator of the truthlikeness of a hypothesis we can use its empirical success, and the problem of the truth conduciveness of explanatory power, consequently, can be analyzed as a problem of confirmation-conduciveness of explanatory power. Based on these considerations we can operationalize claims about truth-conduciveness of IBE. The best explanation is either true or has the highest probability among its competitors. Hence, it has to be either totally successful or be more successful than any of its competitors.

Metaphysical explanatory virtues can only be justified empirically. As Newton-Smith (1981: 224–225) notes, one applies meta-induction (induction on products of inductive reasoning) to discern what features are operative in theories that are considered to be successful. However, this simple association does not suffice to show the truth conduciveness of explanatory virtues. To prove truth conduciveness one has to show that any hypothesis with the ultimate degree of explanatory loveliness has to be successful, and not simply

show that the successful and allegedly true theories exhibit these virtues. Moreover, even if all known hypotheses with the ultimate degree of explanatory power appeared to be successful, this does not mean that IBE is truth-conducive. Inferring the latter is an instance of enumerative induction and this we know to be underdetermined by the evidence. Only a refutation can be conclusive, but successful confirmation of the truth of some conclusion of IBE cannot be conclusive justification of IBE.

Scientific realism is a philosophical view claiming that there are good reasons to believe that well-supported scientific theories are likely to be true or at least approximately true (Kitcher 1993; Leplin 1997; Niiniluoto 1999a; Psillos 1999: e.g.,). Scientific realism explicates truth according to the correspondence theory of truth, according to which, a proposition is true if it corresponds to the state of affairs. Hence, scientific realism claims that theoretical statements of science depict our surrounding reality exactly as it actually is. Scientific realism claims that current scientific theories are true. Scientific realism also claims that current scientific theories are the products of IBE. Therefore, if scientific realism is true, it should follow that the products of IBE are true.

However, the latter conditional is false. One should take into account that there are products of IBE that were or are unsuccessful. The pessimistic induction is one of the most important arguments against scientific realism. Laudan (1981) presented the pessimistic meta-induction argument in order to show that the history of science refutes the claim that the empirical success of particular scientific theories implies the truth of these theories. He gives a list of theories, which in their time were empirically successful, but later become refuted. These theories postulated some theoretical entities that, after theories were refuted, appeared to be nonexistent.

According to Laudan, "a realist would never want to say that a theory was approximately true if its central theoretical terms failed to refer" (Laudan 1981: 33), therefore, the empirical success of scientific theories cannot be the indicator of their truth. A further meta-inductive inference ("meta" here means that it is an inductive inference about the science, which itself is an inductive endeavor, or possible scientific development) concludes that theoretical terms of current empirically successful theories can also fail to refer to actual entities. From the premise that there were empirically successful, but false theories, Laudan makes the conclusion that current empirically successful scientific theories may be false. In other words, empirical success is not sufficient for the truth.

The pessimistic induction is applicable to IBE too. Thagard admits that

> the history of science is replete with highly coherent theories that have turned out to be false, which may suggest that coherence with empirical evidence is a poor guide to truth. (Thagard 2007b: 28)

Newman provides a more particular example of explanatory lovely theories which appeared to be false:

> one could plausibly argue that unification was an important component in the acceptance of both phlogiston and caloric theories. Phlogiston provided a unifying explanation not merely for processes of combustion, but also for calcination, respiration, and smelting. Similarly, caloric explained not merely the transfer of heat, but was also used to derive the adiabatic gas laws, and provided more accurate predictions for the speed of sound in air. (Newman 2009: 126)

Bird provides yet another example:

> good explanations are frequently falsified and often replaced by less virtuous ones. The theory of relativity is less simple than the Newtonian mechanics it replaced, while many aspects of quantum theory are distinctly lacking in virtue and can even be regarded as explanatorily vicious (renormalization, non-locality, complementarity, and so on). The ancient theory of four elements was replaced by one with over one hundred elements. Even if the balance seemed to be restored by the discovery of the three subatomic components of atoms, it was put out of kilter by the subsequent discovery of a zoo of such particles. (Bird 2005: 6)

Hence, all of the above theories were in some respect accepted as best explanations, but turned out to be false. The pessimistic induction implies that IBE is not truth-conducive.

The pessimistic induction is even more damaging to the truth-conduciveness of IBE than to the truthlikeness of scientific realism. For scientific realism to be true, it suffices that current successful scientific theories are true or close to the truth, no matter how they were discovered or accepted. Hence, pessimistic induction will fail to refute scientific realism if current scientific theories are actually true even if some of the past scientific theories were false. However, for IBE to be truth-conducive, it does not suffice that current instances of IBE are true. For IBE to be truth-conducive, all instances of IBE have to be true and if, as the pessimistic induction indicates, there were instances of false conclusions of IBE, then IBE cannot be truth-conducive. Current instances of IBE may be true because the background knowledge is now more accurate and provides a much

smaller set of potential explanations for relevant problems. But being simply the best explanation for a hypothesis does not suffice to be truthlike, as the pessimistic induction indicates. The pessimistic induction thus deductively refutes the truth-conduciveness of IBE.

Even though Thagard (2007b) admits that there were many very coherent theories that have turned out to be false in the history of science he claims that theories that possess certain features were not among the refuted theories and because of that can be considered to be true. More particularly, Thagard notes that theories that both unify and are deepened over time were never rejected as false. He proposes a deepening maxim:

> explanatory coherence leads to truth when a theory not only is the best explanation of the evidence, but also broadens its evidence base over time and is deepened by explanations of why the theory works. (Thagard 2007b: 37)

and

> the deepening maxim can then be specified as the induction that theories can be judged to be true if they have been deepened by having the mechanisms they describe decomposed into more fundamental mechanisms for which there is independent evidence. As we have seen, inductive support for the deepening maxim includes the germ theory of disease, the neuronal theory of brains, molecular cell biology, molecular genetics, and the atomic theory of matter. (Thagard 2007b: 40)

Hence, what Thagard claims is that unification alone is susceptible to the pessimistic induction, but unification and deepening together are not. In other words, Thagard claims that unification alone is not sufficient for the truth-conduciveness of explanatory power and that

deepening is necessary for the truth-conduciveness of explanatory power.

However, the deepening maxim cannot salvage the truth-conduciveness of IBE. Thagard's argument is an argument in support of scientific realism rather than in support of IBE. He admits the falsehood of discarded theories; hence he admits the falsehood of some instances of IBE. If the deepening maxim were to save IBE, it could do it only by arguing that discarded theories were not the best explanations, and that they were lacking in explanatory depth. However, they were accepted as the best at their time, hence, if IBE is truth-conducive, they should have been true. If for some scientists they were the best, but for Thagard they are not, then it only strengthens another argument that IBE cannot be truth-conducive, because there often may be instances of contradicting orderings of explanatory loveliness.

2.6 The Refutation of Truth-Conduciveness of IBE

Neither of the enumerated ways of justification gives a satisfactory justification of the truth-conduciveness of IBE. This section claims that IBE is a genuinely inductive kind of inference, that its inductive features undermine its truth-conduciveness and that its truth-conduciveness is refuted by yet further arguments.

IBE does not conform to the logical rules of valid inference. IBE is an instance of abduction. The logical form of abduction is the following:

$$\{C, A \rightarrow C\} \vdash A.$$

This logical form is a logical fallacy known by the name Affirming the Consequent and sometimes by the name *modus morons* (Thagard

2007b: 288). This logical form is not deductively valid, because the truth of the consequent of the conditional does not entail the truth of the antecedent. For example, suppose we see through the window that the grass is wet. We know that grass gets wet if it rains on it. The probability of the rain as the actual explanation of the wet grass can be further enhanced by the presence of dark clouds and the morning weather forecast that it should rain. Nevertheless, all this does not entail that it actually rained. Maybe somebody watered the grass to deliberately deceive us or maybe an alien spaceship dropped a water bomb on our lawn. These two alternative hypotheses are equally logically possible as the rain hypothesis. One can always present a description of a possible world, in which any conclusion of IBE would be false.

Abduction is the first of the two steps of IBE. One can argue that maybe the second step, which mimics the disjunctive syllogism, makes IBE valid. However, for the disjunctive syllogism to necessitate a true conclusion, the premises of the inference have to be true. In the case of IBE, it means that the set of potential hypotheses has to be exhaustive. We argued (2.2) that in the case of IBE it is impossible to have a set of potential hypotheses that is exhaustive. All the potential hypotheses have to be the conclusions of abductions, hence they will be invalid. A catch-all hypothesis—a negation of all the potential hypotheses in the set—would make the set exhaustive, but it cannot be added to the set of potential hypotheses if it is not explanatory, i.e., if it is not a conclusion of abduction, and a catch-all hypothesis cannot be explanatory in relation to a particular problem at hand, because it explains the negation of the relevant abductive trigger rather than the relevant abductive trigger itself. For example, a blown fuse is an explanation for the failure of the light bulb, but a catch-all for it "The fuse did not blow" will not ex-

plain the failure of the light bulb, because, according to this fact, the light bulb should remain lit. Hence every instance of IBE remains an instance of abduction and, consequently, remains an instance of Affirming the Consequent.

The set of potential explanations cannot be exhaustive due to one more reason. Every instance of IBE is underdetermined by the evidence, because any abductive trigger is consistent with a potential infinity of possible explanations. Any abductive trigger is true. One of the tautologies in the propositional logic $(C \rightarrow (A \rightarrow C))$ states that a true proposition is inferable from any other proposition. Hence, everything that is not proven otherwise is a potential explanation, and everything that is not disproved is a legitimate abductive inference. For example, a simple disjunction introduction (Law of addition) guarantees that alternative explanations, although most often not very plausible, can be generated infinitely. Therefore, one cannot claim to know the best explanation, because one can never evaluate all the possible explanatory hypotheses.

These two reasons strengthen and elaborate the bad lot argument. In every instance of IBE the true explanation might not be among the explanations selected for the evaluation. The theories of IBE are underdetermined by the bad lot argument.

The theories of IBE are further underdetermined by the possibility of competing incompatible rankings of explanatory power (2.2). Explanatory power is evaluated by the multiplicity of criteria, i.e., by the multiplicity of explanatory virtues. Even different aspects of one and the same explanatory virtue can produce contradictory rankings of explanatory power. Contradictory ranking would identify different hypotheses as the best ones, but all of them cannot be true at the same time. Even if we picked a particular hypothesis as the best one, another hypothesis that is better according to some

other explanatory virtues can be the actual explanation. Not only the explanatory hypothesis, but the applied explanatory virtues may also be of a bad lot.

The bad lot argument is further strengthened by the pessimistic induction argument. Pessimistic induction presents examples of lovely theories that are false. That is, there are theories that were the best explanations at their time, but now they are known to be false and are superseded by even better explanations. Hence every example of a superseded theory from the pessimistic induction argument is an example of the bad lot situation in action. Stanford (2006) devoted a whole book to show that scientists repeatedly failed to include in the pool of potential explanations those hypotheses that later scientists accepted as better ones. According to him,

> we have abundant evidence that in past cases we have failed to canvas all of the likely, plausible, or well-confirmed theoretical explanations of the data before proceeding to eliminate alternatives. (Stanford 2006: 31)

On the other hand, the pessimistic induction argument warrants an even stronger conclusion. The pessimistic induction argument applied to the theories of IBE refutes the truth-conduciveness of IBE (2.5.2). Being the best does not suffice to be truthlike. There were instances of false conclusions of IBE; therefore IBE cannot be truth-conducive.

Finally, the better-safe-than-sorry beliefs are pragmatically warranted and useful, albeit false, beliefs. The rationale behind the better-safe-than-sorry beliefs (2.3.3) precludes the possibility of justifying truth-conduciveness of IBE on the grounds that it is pragmatically indispensable.

The theories of IBE set greater requirements for its inferences

than abduction, but inferences of both IBE and abduction are equally truthlike, i.e., IBE, in the same way as abduction and because it is an instance of abduction, provides only potential explanations. Even though the conclusions of IBE are more reasoned than the conclusions of mere abduction, the mechanism of IBE does not grant that the conclusions of IBE are true. More particularly, coherence with background knowledge very strongly constrains the set of potential explanations. Some conclusions may appear to be more surer than others, but this can happen only because background knowledge constrains the set of potential explanations more strongly. When the relevant background knowledge is scarce, IBE should be erroneous very often. When there is a considerable amount of relevant background knowledge, IBE should be erroneous less often. When there is ample relevant background knowledge IBE should be certain or almost certain. Nevertheless, the IBE in all these cases is as equally probable an inference as any other possible inference that is consistent with background knowledge. Therefore, IBE (consistency with background knowledge plus explanatory relations) cannot be said to be truer than mere abduction (mere consistency with background knowledge). IBE provides a more fine grained set of potential explanations; however, the fine graininess is no guarantee that the actual explanation will be an element of this set. IBE only gives reasons to believe in the truth, but is not sufficient to grant it. IBE is a genuine inductive inference with all the inductive flaws.

Many philosophers exhibit confidence about IBE which, for the moment, is unwarranted if one is to scrutinize IBE: current ways of justifying the truth-conduciveness of IBE are unsuccessful. Nevertheless, this conclusion does not imply that there cannot be any other possible way to justify the truth-conduciveness of IBE. A widespread use of IBE is rather successful and this success can be

understood as an abductive trigger that needs to be explained. A promising research project is to scrutinize the concept of material inference. This thesis maintains IBE to be a kind of material inference, i.e., an inference when the truth of the conclusion, given true premises, depends on the meanings of the non-logical terms and, consequently, on the content of propositions in the inference. Hence, if IBE is actually truth-conducive, then there has to operate a peculiar kind of validity, which may be called "material validity." The concept of material inference is rather rarely mentioned in the philosophical literature. Material inference is often associated with Sellars (1953), but the concept can be traced back at least to the works of Jean Buridan. Only lately it began to gain slightly greater interest among philosophers (e.g., Brandom 1994; 2001; Brigandt 2010; Norton 2003; Read 1994).

To conclude, there are currently no satisfactory conceptual, formal or historical reasons why IBE has to be truth-conducive. The belief in the truth of best explanations is unjustified, even if we have a cognitive bias to infer this kind of hypotheses and even if some hypotheses later appear to be true. Conclusions of IBE are true contingently rather than inherently. The most that the theories of IBE can argue is not that IBE is truth-conducive, but, as the evolutionary psychology interpretation of IBE as a psychological fact suggests, IBE provides true conclusions more often than any other known ampliative method of reasoning. However, if this is true, it does not imply that conclusions of IBE are true more often than false. It provides an easier or more convenient way to seek the truth than other forms of ampliative inference, but it does not always lead to the truth. It is a convenient and psychologically compelling way to assign probability distributions, but empirical conditionalization would most often correct and adjust the assigned probability dis-

tribution for a hypothesis. Hence pragmatic warrant remains the strongest justification of IBE: currently there is no better method of ampliative inference. However, being pragmatically warranted does not make IBE truth-conducive.

IBE is in its essence an instance of the *ad Ignorantiam* (the argument from ignorance) fallacy: acceptance of a conclusion on the basis that it has not yet been proven false. This is most evident in the quote of Psillos

> if a hypothesis has been chosen as the best explanation, then it has fared best in an explanatory-quality test with its competing rivals. So unless there is reason to think that it is superseded by an even better explanation, or unless there is reason to believe that the recalcitrant evidence points to one of the rivals as a better explanation, to stick with the best explanatory hypothesis is entirely reasonable. (Psillos 2002: 622)

Thus logically IBE is a fallacy, but in ordinary life it is a mode of operation.

Conclusions

1. IBE is a form of material inference that ascribes truth to the hypothesis that has the highest degree of explanatory virtues among its competitors: it is the most consistent with approved background knowledge, the most unifying, the deepest and the simplest. If a hypothesis is not consistent with or does not restore consistency in the relevant background knowledge, then the hypothesis cannot be an abductive conclusion and, because of that, cannot be included in the set of potential explanations. In addition to consistency unification as an explanation of several different kinds of phenomena, explanatory depth as an explanation of why the proposed explanation should be true, and simplicity as fewness of posited explanatory entities or shortness of expression of an explanation are features attributed to explanations having a sufficient rationale to be accepted. Therefore, if explanations exhibiting these features are good explanations, then explanations exhibiting the highest degree of consistency, unification, depth and simplicity are the best explanations. Nevertheless, there are two inconsistencies in the theories of IBE:

   (a) The theories that treat coherence as the main explanatory virtue are circular, because coherence itself stands for consistency plus explanatory relations. Therefore, explanatory power and coherence should be considered to stand for the

very same phenomenon and explicated as derivable from the rest of explanatory virtues.

(b) Probabilistic measures of explanatory power are superfluous: if probability distributions are known, then the posterior probability formula is the correct way to identify the most likely hypothesis, but this ceases to be IBE; if probability distributions are not known, then one cannot apply probabilistic measures of explanatory power and has to rely on the explanationist account of explanatory power. Either way, there is no actual need for the probabilistic measures of the explanatory power. The explanationist account of explanatory power is sufficient for the task at hand.

2. Current theories of IBE can be classified into four basic ways of justifying the truth-conduciveness of IBE:

   (a) The reliabilist-coherentist way claims that IBE is truth-conducive, because it is reliable—prone to produce true beliefs rather than false ones—and it is reliable, because it enhances the total coherence of knowledge.

   (b) The evolutionary way claims that IBE is truth-conducive, because reasoning following IBE is a survival-enhancing human cognitive capacity and the best explanation for why IBE is survival-enhancing is because IBE provides true beliefs.

   (c) The probabilistic way claims that IBE can be defined probabilistically and IBE is truth-conducive, because hypotheses with the highest degree of explanatory virtues appear to have the highest posterior probability.

   (d) The empirical-historical way claims that IBE is truth-conducive, because hypotheses with the highest degree of ex-

planatory virtues are empirically confirmed and truth or at least approximate truth is the best explanation for the empirical confirmation.

3. None of the four ways of justification discerned above grants the truth-conduciveness of IBE:

   (a) Better-safe-than-sorry argument refutes the evolutionary way of justification: there are survival-enhancing beliefs that do not have to be true.

   (b) Pessimistic induction refutes the empirical-historical way of justification: there were instances of best explanations that were accepted as true, but later appeared to be false.

   (c) The possibility of contradicting orders of explanatory power undermines the reliabilist-coherentist way of justification: one can never be sure that there is no alternative, and more adequate, ranking of explanatory power.

   (d) The bad lot argument undermines the reliabilist-coherentist way of justifying IBE: the set of potential explanations can never be exhaustive and the use of catch-all hypotheses cannot correct that. Pessimistic induction strengthens the bad lot argument: every example of a superseded theory from the pessimistic induction argument is an example of the bad lot situation in action; a better theory, which superseded the old one, was not considered when the old one was accepted. The possibility of contradicting orders of explanatory power strengthens the bad lot argument: even if we picked a particular hypothesis as the best one, another hypothesis that is better according to some other explanatory virtues can be the actual explanation.

(e) The material nature of IBE contradicts an application of the probabilistic way to argue for the truth-conduciveness of IBE. Probabilistic justification would make IBE deductively, i.e., formally valid. However, IBE is a non-deductive kind of inference; its alleged validity is not formal, but material, dependent on the substantive considerations.

(f) Further research on the concept of material inference has to be conducted in order to assess whether there is any way to show that IBE is truth-conducive.

4. Even though, for the time being, IBE cannot be showed to be truth-conducive, IBE is a widespread psychological practice to provide hypotheses that can later be empirically tested. After successful empirical tests, conclusions of IBE can be given credence, but not before. Therefore, IBE is warranted at most pragmatically: there is currently no better kind of ampliative inference and its application helps us to successfully cope with the world. However, pragmatic warranty does not imply the truth-conduciveness of IBE.

Bibliography

Ahn, W., Kalish, C. W., Medin, D. L. and Gelman, S. A. (1995). The Role of Covariation Versus Mechanism Information in Causal Attribution. *Cognition* 54(3): 299–352. doi:10.1016/0010-0277(94)00640-7.

Aliseda, A. (1997). *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence*. Ph.D. thesis, Stanford University.

Aliseda, A. (2006). *Abductive Reasoning*, *Synthese Library*, vol. 330. Dordrecht: Springer. doi:10.1007/1-4020-3907-7.

Almeder, R. (2007). Pragmatism and Philosophy of Science: A Critical Survey. *International Studies in the Philosophy of Science* 21(2): 171–195. doi:10.1080/02698590701498100.

Barnes, E. (1995). Inference to the Loveliest Explanation. *Synthese* 103(2): 251–227. doi:10.1007/BF01090049.

Bartelborth, T. (1999). Coherence and Explanations. *Erkenntnis* 50(2-3): 209–224. doi:10.1023/A:1005594409663.

Bartelborth, T. (2002). Explanatory Unification. *Synthese* 130(1): 91–108. doi:10.1023/A:1013827209894.

Bartelborth, T. (2005). Is the Best Explaining Theory the Most Probable One? *Grazer Philosophische Studien* 70: 1–23.

Beebe, J. R. (2009). The Abductivist Reply to Skepticism. *Philosophy and Phenomenological Research* 79(3): 605–636. doi:10.1111/j.1933-1592.2009.00295.x.

Ben-Menahem, Y. (1990). The Inference to the Best Explanation. *Erkenntnis* 33(3): 319–344. doi:10.1007/BF00717590.

Bird, A. (2005). Abductive Knowledge and Holmesian Inference. In T. S. Gendler and J. Hawthorne (Eds.) *Oxford Studies in Epistemology: Volume 1*, New York: Oxford University Press. pp. 1–31.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

Boulter, S. J. (2007). The "Evolutionary Argument" and the Metaphilosophy of Commonsense. *Biology and Philosophy* 22(3): 369–382. doi:10.1007/s10539-006-9032-z.

Bovens, L. and Hartmann, S. (2003). *Bayesian Epistemology*. Oxford: Clarendon Press.

Bovens, L. and Olsson, E. J. (2000). Coherentism, Reliability and Bayesian Networks. *Mind* 109(436): 685–719. doi:10.1093/mind/109.436.685.

Bovens, L. and Olsson, E. J. (2002). Believing More, Risking Less: On Coherence, Truth and Non-Trivial Extensions. *Erkenntnis* 57(2): 137–150. doi:10.1023/A:1020913625002.

Brandom, R. B. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.

Brandom, R. B. (2001). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.

Brem, S. K. and Rips, L. J. (2000). Explanation and Evidence in Informal Argument. *Cognitive Science* 24(4): 573–604. doi:10.1016/S0364-0213(00)00033-1.

Brigandt, I. (2010). Scientific Reasoning is Material Inference: Combining Confirmation, Discovery, and Explanation. *International Studies in the Philosophy of Science* 24(1): 31–43. doi:10.1080/02698590903467101.

Campos, D. G. (2009). On the Distinction between Peirce's Abduction and Lipton's Inference to the Best Explanation. *Synthese* 180(3): 419–442. doi:10.1007/s11229-009-9709-3.

Carrier, M. (2009). Underdetermination as an Epistemological Test Tube: Expounding Hidden Values of the Scientific Community. *Synthese* 180(2): 189–204. doi:10.1007/s11229-009-9597-6.

Carruthers, P. (1992). *Human Knowledge and Human Nature*. New York: Oxford University Press.

Carruthers, P. (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. New York: Oxford University Press.

Collins, H. M. and Pinch, T. (1993). *The Golem: What Everyone Should Know about Science*. Cambridge: Cambridge University Press.

Day, T. and Kincaid, H. (1994). Putting Inference to the Best Explanation in its Place. *Synthese* 98(2): 271–295. doi:10.1007/BF01063944.

de Finetti, B. (1964). Foresight: Its Logical Laws, Its Subjective Sources. In H. E. J. Kyburg and H. E. Smokler (Eds.) *Studies in Subjective Probability*, New York: John Wiley & Sons. pp. 93–158.

De Smedt, J. and De Cruz, H. (2010). Evolved Cognitive Biases and the Epistemic Status of Science. In *Epistemology and Philosophy of Mind at The Crossroads*. Institute of Philosophy, University of Leuven: Fourth Conference of the Dutch-Flemish Association for Analytic Philosophy (VAF IV), January 20–22, 2010.

Douven, I. (1999). Inference to the Best Explanation Made Coherent. *Philosophy of Science* 66(Supplement): S424–S435. doi:10.1086/392743.

Douven, I. (2002). Testing Inference to the Best Explanation. *Synthese* 130(3): 355–377. doi:10.1023/A:1014859910339.

Douven, I. (2011). Abduction. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Spring 2011 ed.

Douven, I. and Horsten, L. (1998). Earman on Underdetermination and Empirical Indistinguishability. *Erkenntnis* 49(3): 303–320. doi:10.1023/A:1005437217700.

Enoch, D. and Schechter, J. (2008). How are Basic Belief-Forming Methods Justified? *Philosophy and Phenomenological Research* 76(3): 547–579. doi:10.1111/j.1933-1592.2008.00157.x.

Fitelson, B. (2003). A Probabilistic Theory of Coherence. *Analysis* 63(3): 194–199. doi:10.1111/1467-8284.00420.

Floridi, L. (2009). Logical Fallacies as Informational Shortcuts. *Synthese* 167(2): 317–325. doi:10.1007/s11229-008-9410-y.

Fodor, J. A. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*. Cambridge, MA: MIT Press.

Forster, M. and Sober, E. (1994). How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions. *The British Journal for the Philosophy of Science* 45(1): 1–35. doi:10.1093/bjps/45.1.1.

Gabbay, D. M. and Woods, J. (2005). *The Reach of Abduction: Insight and Trial, A Practical Logic of Cognitive Systems*, vol. 2. Amsterdam: Elsevier.

Glass, D. H. (2002). Coherence, Explanation, and Bayesian Networks. In M. O'Neill, R. F. E. Sutcliffe, C. Ryan, M. Eaton and N. J. L. Griffith (Eds.) *Artificial Intelligence and Cognitive Science*, Berlin-Heidelberg: Springer, *Lecture Notes in Artificial Intelligence*, vol. 2464. pp. 35–145. doi: 10.1007/3-540-45750-X_23.

Glass, D. H. (2007). Coherence Measures and Inference to the Best Explanation. *Synthese* 157(3): 275–296. doi:10.1007/s11229-006-9055-7.

Glass, D. H. (2010). Inference to the Best Explanation: Does It Track Truth? *Synthese* doi:10.1007/s11229-010-9829-9.

Goldman, A. H. (1990). Natural Selection, Justification, and Inference to the Best Explanation. In N. Rescher (Ed.) *Evolution, Cognition, and Realism: Studies in Evolutionary Epistemology*, CPS Series in Philosophy of Science, Lanham, MD: University Press of America. pp. 39–46.

Grünbaum, A. (2007). Is Simplicity Evidence of Truth? *Royal Institute of Philosophy* 82(Supplement 61): 261–275. doi:10.1017/S1358246107000227.

Harman, G. H. (1965). The Inference to the Best Explanation. *The Philosohical Review* 74(1): 88–95. doi:10.2307/2183532.

Harman, G. H. (1968). Enumerative Induction as Inference to the Best Explanation. *The Journal of Philosophy* 65(18): 529–533. doi:10.2307/2024115.

Harman, G. H. (1970). Induction. A Discussion of the Relevance of the Theory of Knowledge to the Theory of Induction (with a Digression to the Effect that neither Deductive Logic nor the Probability Calculus has Anything to Do with Inference). In M. Swain (Ed.) *Induction, Acceptance, and Rational Belief*, Synthese Library, Dordrecht: D. Reidel Publishing Company. pp. 83–99.

Harman, G. H. (1999). *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press. doi:10.1093/0198238029.001.0001.

Harris, S., Sheth, S. A. and Cohen, M. S. (2008). Functional Neuroimaging of Belief, Disbelief, and Uncertainty. *Annals of Neurology* 63(2): 141–147. doi: 10.1002/ana.21301.

Hintikka, J. (1998). What is Abduction? The Fundamental Problem of Contemporary Epistemology. *Transactions of the Charles S. Peirce Society* 34(3): 503–533.

Hitchcock, C. (2007). The Lovely and the Probable. *Philosophy and Phenomenological Research* 74(2): 433–440. doi:10.1111/j.1933-1592.2007.00029.x.

Hon, G. and Rakover, S. S. (Eds.) (2001). *Explanation: Theoretical Approaches and Applications*. Dordrecht: Kluwer Academic Publishers.

Huemer, M. (2009a). Explanationist Aid for the Theory of Inductive Logic. *The British Journal for the Philosophy of Science* 60(2): 345–375. doi:10.1093/bjps/axp008.

Huemer, M. (2009b). When is Parsimony a Virtue? *The Philosophical Quarterly* 59(235): 216–236. doi:10.1111/j.1467-9213.2008.569.x.

Humphreys, P. (1993). Greater Unification Equals Greater Understanding? *Analysis* 53(3): 183–188. doi:10.2307/3328470.

Josephson, J. R. (2001). On the Proof Dynamics of Inference to the Best Explanation. *Cardozo Law Review* 22(5): 1621–1643.

Josephson, J. R. and Josephson, S. G. (Eds.) (2003). *Abductive inference: Computation, Philosophy, Technology*. Cambridge: Cambridge University Press.

Kamps, J. (2005). The Ubiquity of Background Knowledge. In R. Festa, A. Aliseda and J. Peijnenburg (Eds.) *Cognitive Structures in Scientific Inquiry. Essays in Debate with Theo Kuipers Volume 2*, Amsterdam: Rodopi, *Poznan Studies in the Philosophy of the Sciences and the Humanities*, vol. 84. pp. 317–337.

Kelly, K. T. (2007). A New Solution to the Puzzle of Simplicity. *Philosophy of Science* 74(5): 561–573. doi:10.1086/525604.

Kitcher, P. (1993). *The Advancement of Science*. Oxford: Oxford University Press.

Klein, P. and Warfield, T. A. (1994). What Price Coherence? *Analysis* 54(3): 129–132. doi:10.2307/3328660.

Knauff, M. (2007). How Our Brains Reason Logically. *Topoi* 26(1): 19–36. doi:10.1007/s11245-006-9002-8.

Koehler, D. J. (1991). Explanation, Imagination, and Confidence in Judgment. *Psychological Bulletin* 110(3): 499–519. doi:10.1037/0033-2909.110.3.499.

Koslowski, B., Marasia, J., Chelenza, M. and Dublin, R. (2008). Information Becomes Evidence When an Explanation Can Incorporate It Into a Causal Framework. *Cognitive Development* 23(4): 472–487. doi:10.1016/j.cogdev.2008.09.007.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press, third ed.

Kuipers, T. A. F. (2002). Beauty, A Road to the Truth. *Synthese* 131(3): 291–328. doi:10.1023/A:1016188509393.

Kuipers, T. A. F. (2004). Inference to the Best Theory, rather than Inference to the Best Explanation. Kinds of Abduction and Induction. In Stadler (2004), pp. 25–51.

Ladyman, J., Douven, I., Horsten, L. and van Fraassen, B. (1997). A Defence of Van Fraassen's Critique of Abductive Inference: Reply to psillos. *Philosophical Quarterly* 47(188): 305–321. doi:10.1007/BF01063946.

Lakatos, I. (1970). History of Science and Its Rational Reconstructions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1970: 91–136.

Lange, M. (2004). Bayesianism and Unification: A Reply to Wayne Myrvold. *Philosophy of Science* 71(2): 205–215. doi:10.1086/383012.

Latour, B. and Woolgar, S. (1986). *Laboratory Life: The Social Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Laudan, L. (1981). A Confutation of Convergent Realism. *Philosophy of Science* 48(1): 19–49.

Lehrer, K. (1990). *Theory of Knowledge*. London: Routledge.

Leplin, J. (1997). *A Novel Defense of Scientific Realism*. Oxford: Oxford University Press.

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. LaSalle, Ill.: Open Court.

Lipton, P. (1993). Is the Best Good Enough? *Proceedings of the Aristotelian Society* 43: 89–104.

Lipton, P. (2001a). Is Explanation a Guide to Inference? A Reply to Wesley C. Salmon. In Hon and Rakover (2001), pp. 93–120.

Lipton, P. (2001b). What Good is an Explanation? In Hon and Rakover (2001), pp. 53–59.

Lipton, P. (2004). *Inference to the Best Explanation*. London: Routledge, second ed.

Lipton, P. (2007). Replies. *Philosophy and Phenomenological Research* 74(2): 449–462. doi:10.1111/j.1933-1592.2007.00031.x.

Lombrozo, T. (2007). Simplicity and Probability in Causal Explanation. *Cognitive Psychology* 55(3): 232–257. doi:10.1016/j.cogpsych.2006.09.006.

Lycan, W. G. (1988). *Judgement and Justification*. Cambridge: Cambridge University Press.

Magnani, L. (2001). *Abduction, Reason, and Science: Processes of Discovery and Explanation*. New York: Kluwer Academic/Plenum Publishers.

Marques, J. F., Canessa, N. and Cappa, S. (2009). Neural Differences in the Processing of True and False Sentences: Insights into the Nature of 'Truth' in Language Comprehension. *Cortex* 45(6): 759–768. doi:10.1016/j.cortex.2008.07.004.

McAllister, J. W. (1989). Truth and Beauty in Scientific Reason. *Synthese* 78(1): 25–51. doi:10.1007/BF00869680.

McAllister, J. W. (1991). The Simplicity of Theories: Its Degree and Form. *Journal for General Philosophy of Science* 22(1): 1–14. doi:10.1007/BF01801246.

McGrew, T. (2003). Confirmation, Heuristics, and Explanatory Reasoning. *The British Journal for the Philosophy of Science* 54(4): 553–567. doi:10.1093/bjps/54.4.553.

McKaughan, D. J. (2008). From Ugly Duckling to Swan: C. S. Peirce, Abduction, and the Pursuit of Scientific Theories. *Transactions of the Charles S. Peirce Society* 44(3): 446–468.

McMullin, E. (1996). Epistemic Virtue and Theory Appraisal. In I. Douven and L. Horsten (Eds.) *Realism in the Sciences*, Leuven: Leuven University Press. pp. 13–34.

Meijs, W. and Douven, I. (2007). On the Alleged Impossibility of Coherence. *Synthese* 157(3): 347–360. doi:10.1007/s11229-006-9060-x.

Merricks, T. (1995). On Behalf of the Coherentist. *Analysis* 55(4): 306–309. doi:10.2307/3328404.

Minnameier, G. (2004). Peirce-Suit of Truth – Why Inference to the Best Explanation and Abduction Ought Not to be Confused. *Erkenntnis* 60(1): 75–105. doi:10.1023/B:ERKE.0000005162.52052.7f.

Myrvold, W. C. (2003). A Bayesian Account of the Virtue of Unification. *Philosophy of Science* 70(2): 399–423. doi:10.1086/375475.

Newman, M. (2009). The No-Miracles Argument, Reliabilism, and a Methodological Version of the Generality Problem. *Synthese* 177(1): 111–138. doi:10.1007/s11229-009-9642-5.

Newton-Smith, W. H. (1981). *The Rationality of Science*. London: Routledge and Kegan Paul.

Niiniluoto, I. (1999a). *Critical Scientific Realism*. Oxford: Oxford University Press. doi:10.1093/0199251614.001.0001.

Niiniluoto, I. (1999b). Defending Abduction. *Philosophy of Science* 66(Supplement): S436–S451. doi:10.1086/392744.

Niiniluoto, I. (2004). Truth-Seeking by Abduction. In Stadler (2004), pp. 57–82.

Norton, J. D. (2003). A Material Theory of Induction. *Philosophy of Science* 70(4): 647–670. doi:10.1086/378858.

Okasha, S. (2000). Van Fraassen's Critique of Inference to the Best Explanation. *Studies in History and Philosophy of Science* 31(4): 691–710. doi:10.1016/S0039-3681(00)00016-9.

Olsson, E. J. (2002). What is the Problem of Coherence and Truth? *The Journal of Philosophy* 99(5): 246–272. doi:10.2307/3655648.

Olsson, E. J. (2005a). *Against Coherence: Truth, Probability and Justification*. Oxford: Oxford University Press.

Olsson, E. J. (2005b). The Impossibility of Coherence. *Erkenntnis* 63(3): 387–412. doi:10.1007/s10670-005-4007-z.

Paavola, S. (2005). Peircean Abduction: Instinct or Inference? *Semiotica* 153(1/4): 131–154. doi:10.1515/semi.2005.2005.153-1-4.131.

Paavola, S. (2006a). Hansonian and Harmanian Abduction as Models of Discovery. *International Studies in the Philosophy of Science* 20(1): 93–108. doi:10.1080/02698590600641065.

Paavola, S. (2006b). *On the Origin of Ideas: An Abductivist Approach to Discovery*. Ph.D. thesis, University of Helsinki.

Peirce, C. S. (1932). *The Collected Papers of Charles Sanders Peirce*, vol. 2. Cambridge, MA: Harvard University Press.

Peirce, C. S. (1934). *The Collected Papers of Charles Sanders Peirce*, vol. 5. Cambridge, MA: Harvard University Press.

Peirce, C. S. (1935). *The Collected Papers of Charles Sanders Peirce*, vol. 6. Cambridge, MA: Harvard University Press.

Peirce, C. S. (1958). *The Collected Papers of Charles Sanders Peirce*, vol. 7. Cambridge, MA: Harvard University Press.

Persson, J. (2007). IBE and EBI: On Explanation Before Inference. In J. Persson and P. Ylikoski (Eds.) *Rethinking Explanation*, Dordrecht: Springer, *Boston Studies in the Philosophy of Science*, vol. 252. pp. 137–147. doi:10.1007/978-1-4020-5581-2_10.

Plečkaitis, R. (2006). *Logikos pagrindai*. Vilnius: Tyto alba.

Popper, K. R. (2002/1959). *The Logic of Scientific Discovery*. London: Routledge.

Psillos, S. (1996). On Van Fraassen's Critique of Abductive Reasoning. *The Philosophical Quarterly* 46(182): 31–47. doi:10.2307/2956303.

Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. London: Routledge.

Psillos, S. (2000). Abduction: Between Conceptual Richness and Computational Complexity. In A. Flach, Peter and C. Kakas, Antonis (Eds.) *Abduction and Induction: Essays in their Relation and Integration*, Dordrecht: Kluwer Academic Publishers, *Applied Logic Series*, vol. 18. pp. 59–74.

Psillos, S. (2002). Simply the Best: A Case for Abduction. In A. C. Kakas and F. Sadri (Eds.) *Computational Logic: Logic Programming and Beyond*, Berlin-Heidelberg: Springer, *Lecture Notes in Computer Science*, vol. 2408. pp. 605–625. doi:10.1007/3-540-45632-5.

Psillos, S. (2004). Inference to the Best Explanation and Bayesianism. In Stadler (2004), pp. 83–91.

Psillos, S. (2007). The Fine Structure of Inference to the Best Explanation. *Philosophy and Phenomenological Research* 74(2): 441–448. doi:10.1111/j.1933-1592.2007.00030.x.

Psillos, S. (2009a). An Explorer upon Untrodden Ground: Peirce on Abduction. In D. M. Gabbay, S. Hartmann and J. Woods (Eds.) *Handbook of the History and Philosophy of Logic: Inductive Logic*, Amsterdam: Elsevier, vol. 10. pp. 117–151.

Psillos, S. (2009b). Inference to the Best Explanation and Bayesianism. In Psillos (2009c), pp. 195–201.

Psillos, S. (2009c). *Knowing the Structure of Nature: Essays on Realism and Explanation*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

Psillos, S. (2009d). Simply the Best: A Case for Abduction. In Psillos (2009c), pp. 173–194.

Putnam, H. (1975). *Mathematics, Matter and Method: Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press.

Quine, W. V. O. (1969). *Ontological Relativity and Other Essays*. New York: Columbia University Press.

Ramsey, F. P. (1931). Truth and Probability. In R. B. Braithwaite (Ed.) *The Foundations of Mathematics and other Logical Essays*, London: Routledge and Kegan Paul. pp. 156–198.

Read, S. (1994). Formal and Material Consequence. *Journal of Philosophical Logic* 23(3): 247–265. doi:10.1007/BF01048482.

Read, S. J. and Marcus-Newhall, A. (1993). Explanatory Coherence in Social Explanations: A Parallel Distributed Processing Account. *Journal of Personality and Social Psychology* 65(3): 429–447. doi:10.1037/0022-3514.65.3.429.

Reichenbach, H. (1956). *The Direction of Time*. Los Angeles, CA: University of California Press.

Rescher, N. (1973). *The Coherence Theory of Truth*. Oxford: Clarendon Press.

Richter, T., Schroeder, S. and Woehrmann, B. (2009). You Don't Have to Believe Everything You Read: Background Knowledge Permits Fast and Efficient Validation of Information. *Journal of Personality and Social Psychology* 96(3): 538–558. doi:10.1037/a0014038.

Salmon, W. C. (1970). Bayes's Theorem and the History of Science. In R. H. Stuewer (Ed.) *Historical and Philosophical Perspectives of Science*, Minneapolis, MN: University of Minnesota Press, *Minnesota Studies in the Philosophy of Science*, vol. 5. pp. 68–86.

Salmon, W. C. (1989). *Four Decades of Scientific Explanation*. Pittsburgh, PA: University of Pittsburgh Press.

Salmon, W. C. (1990). Rationality and Objectivity in Science, or Tom Kuhn Meets Tom Bayes. In C. W. Savage (Ed.) *Scientific Theories*, Minneapolis, MN: University of Minnesota Press, *Minnesota Studies in the Philosophy of Science*, vol. 14. pp. 175–204.

Salmon, W. C. (2001a). Explanation and Confirmation: A Bayesian Critique of Inference to the Best Explanation. In Hon and Rakover (2001), pp. 61–91.

Salmon, W. C. (2001b). Reflections of A Bashful Bayesian: A Reply to Lipton. In Hon and Rakover (2001), pp. 121–136.

Schupbach, J. N. (2005). On a Bayesian Analysis of the Virtue of Unification. *Philosophy of Science* 72(4): 594–607. doi:10.1086/505186.

Schupbach, J. N. and Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science* 78(1): 105–127. doi:10.1086/658111.

Schurz, G. (1999). Explanation as Unification. *Synthese* 120(1): 95–114. doi:10.1023/A:1005214721929.

Schurz, G. (2008). Patterns of Abduction. *Synthese* 164(2): 201–234. doi:10.1007/s11229-007-9223-4.

Sellars, W. (1953). Inference and Meaning. *Mind* 62(247): 313–338. doi:10.1093/mind/LXII.247.313.

Shogenji, T. (1999). Is Coherence Truth Conducive? *Analysis* 59(4): 338–345. doi:10.1111/1467-8284.00191.

Simon, A. H. (2001). Science Seeks Parsimony, Not Simplicity: Searching for Pattern in Phenomena. In Zellner et al. (2001), pp. 32–72.

Sober, E. (1990). Let's razor Ockham's Razor. In D. Knowles (Ed.) *Explanation and its Limits*, Cambridge: Cambridge University Press. pp. 73–93.

Sober, E. (2001). What is the Problem of Simplicity? In Zellner et al. (2001), pp. 13–31.

Sober, E. (2002). Bayesianism – its Scope and Limits. In R. Swinburne (Ed.) *Bayes's Theorem*, Oxford: Oxford University Press, *Proceedings of the British Academy*, vol. 113. pp. 21–38.

Stadler, F. (Ed.) (2004). *Induction and Deduction in the Sciences*, *Vienna Circle Institute Yearbook*, vol. 11. Dordrecht: Kluwer.

Stanford, K. P. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. New York: Oxford University Press.

Stephens, C. L. (2001). When is it Selectively Advantageous to Have True Beliefs? Sandwiching the Better Safe than Sorry Argument. *Philosophical Studies* 105(2): 161–189. doi:10.1023/A:1010358100423.

Stich, S. P. (1990). *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation*. Cambridge, MA: MIT Press.

Teller, P. (1973). Conditionalization and Observation. *Synthese* 26(2): 218–258. doi:10.1007/BF00873264.

Thagard, P. (1978). The Best Explanation: Criteria for Theory Choice. *The Journal of Philosophy* 75(2): 76–92. doi:10.2307/2025686.

Thagard, P. (1989). Explanatory Coherence. *Behavioural and Brain Sciences* 12: 435–502.

Thagard, P. (1993). *Computational Philosophy of Science*. Cambridge, MA: MIT Press.

Thagard, P. (2007a). Abductive Inference: From Philosophical Analysis to Neural Mechanisms. In A. Feeney and E. Heit (Eds.) *Inductive Reasoning: Experimental, Developmental, and Computational Approaches*, Cambridge: Cambridge University Press. pp. 226–247.

Thagard, P. (2007b). Coherence, Truth, and the Development of Scientific Knowledge. *Philosophy of Science* 74(1): 28–47. doi:10.1086/520941.

Tuomela, R. (1985). Truth and Best Explanation. *Erkenntnis* 22(1–3): 271–299. doi:10.1007/BF00269971.

Tversky, A. and Kahneman, D. (1982). Judgments of and by Representativeness. In *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press. pp. 84–98.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford: Oxford University Press. doi:10.1093/0198244274.001.0001.

van Fraassen, B. C. (1985). Empiricism in the Philosophy of Science. In P. M. Churchland and C. A. Hooker (Eds.) *Images of Science: Essays on Realism and Empiricism*, Chicago, IL: University of Chicago Press. pp. 245–308.

van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford: Oxford University Press. doi:10.1093/0198248601.001.0001.

Weber, M. (2009). The Crux of Crucial Experiments: Duhem's Problems and Inference to the Best Explanation. *The British Journal for the Philosophy of Science* 60(1): 19–49. doi:10.1093/bjps/axn040.

Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese* 167(1): 125–143. doi:10.1007/s11229-008-9305-y.

Ylikoski, P. and Kuorikoski, J. (2010). Dissecting Explanatory Power. *Philosophical Studies* 148(2): 201–219. doi:10.1007/s11098-008-9324-z.

Zellner, A., Keuzenkamp, H. A. and McAleer, M. (Eds.) (2001). *Simplicity, Inference and Modelling: Keeping it Sophisticatedly Simple*. Cambridge: Cambridge University Press.