Pažinimo ir mokslo filosofija

PSYCHOLOGICAL ADEQUACY AND ONTOLOGICAL COMMITMENTS OF INFERENCE TO THE BEST EXPLANATION

Adolfas Mackonis
Vilniaus universiteto
Filosofijos istorijos ir logikos katedra
Universiteto g. 9/1, LT-01513 Vilnius
Tel. (8 5) 266 7617
El. paštas adolfas.mackonis@fsf.stud.vu.lt

The article explicates psychological and ontological aspects of Inference to the Best Explanation (IBE). IBE is a psychological theory, because cognitive science studies support IBE as descriptively true and psychologically adequate theory, i.e., people perceive best explanations as true and follow the rule of IBE in their reasoning. Moreover, different features of IBE imply that conclusions of IBE can be true only in a world with a very particular ontological constitution. Realism about the external world, the uniformity of nature, the truth of background knowledge and the truth-conduciveness of explanatory virtues are necessary and sufficient for IBE to be truth-conducive. Therefore, IBE is an epistemic theory only because at the same time it is committed to a particular ontology.

***Keywords*:** *inference to the best explanation, abduction, psychology, ontology, explanatory virtues.*

The Psychological Hypothesis

Inference to the best explanation (IBE) is an epistemic theory about the power of explanatory considerations. According to the theory of IBE, if one wants to know what is the cause of some surprising fact one should infer the best explanation for it and this explanation will be true. The explanation will depict the actual state of affairs that caused the surprising fact. These claims are not uncontroversial and are widely discussed in the analytical epistemology and philosophy of science literature. Sometimes studies in epistemology and philosophy of science that discuss IBE are concerned with IBE as a descriptive theory. When this happens these studies analyze whether scientists actually rely on IBE when choosing between theories. However, these studies in epistemology and philosophy of science never 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)

Similarly, Gabbay and Woods state about abduction, which is very often associated with IBE:

> [...] for any abduction problem, even if it is justified to postulate the existence of a filtration structure in which abductive solutions are cutdowns of up to very large possibility spaces, there is no empirical evidence that real-life abducers achieve their abductive targets by constructing such structures. (Gabbay & Woods 2005: 155)

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 psychological 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 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 to 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 idea that IBE is a human cognitive capacity is not recent. Theory of IBE merely develops Peirce's claims about abduction or, if the distinction between early and late Peirce (Gabbay & Woods 2005: 40; Psillos 2009: 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 the pure chance would let it (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).

A critical reader can notice the features of IBE in the argumentation for the psychological hypothesis. The psychological hypothesis is coherent with experimental findings and its truth would explain why invalid and underdetermined form of inference appears for some philosophers a justified and reliable one. This is circular, i.e., IBE is used to establish the psychological adequacy of IBE. However, if the psychological hypothesis is true, people, including the author of this article, simply cannot escape the use of IBE. IBE might be the only ampliative form of inference that people accept or perceive as reliable.

IBE is a strengthened form of abduction. IBE adds one premise to the abductive

form of inference, addition of which, supposedly, permits to infer not merely a possible, but true conclusion. Peirce presented a canonical example of the abductive form of inference:

> 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 adds to this form of inference an additional premise that there is no better explanation for the surprising fact than the hypothesis analyzed (Harman 1965: 89; Psillos 2002: 614). The addition of this premise strengthens the conclusion. If the conclusion of abduction "merely suggests that something may be" (Peirce 1934: 5.171) or is "worthy of conjecture" (Gabbay & Woods 2005: 69), the conclusion of IBE is "true" (Harman 1965: 89), "true, or at least approximately true" (Lipton 2004: 3) or "probably true" (Josephson & Josephson 2003: 5; Psillos 2002: 614). The word "probably" in the last quote is used without any specific interpretation of probability theory, it merely acknowledges that the conclusion is not a deductive one. Thus the form of IBE is this:

> 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,
>
> Hence, A is true.

The logical form of IBE is abductive. Every instance of IBE is an instance of abduction, but not every instance of abduction is an instance of IBE. Both abduction and IBE are non-deductive forms of arguments, because the logical form of abduction and, consequently, of IBE is deductively invalid. It is very easy to imagine a state of affairs that would make the premises of abduction true, but where the conclusion would be false. For example, suppose the light bulb in a room unexpectedly goes out. One knows that this can happen due to the fuse, i.e. the blown safety fuse would explain the failure of the light bulb in the room. However, it is also completely compatible with an accident in the power plant or even an airplane crash that tore off the power line that supplied electricity to the house. IBE chooses the best explanation in the set of the abductive, i.e. deductively invalid, conclusions. Whichever abductive conclusion is chosen as the best explanation, it remains deductively invalid, because it remains abductive. Hence, IBE inherits deductive invalidity of abduction. One might argue that an addition of an 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, in order to be accepted as an abductive conclusion, the catch-all hypothesis has to be a genuine explanation and catch-all hypotheses are not genuine explanations, not to say good enough explanations. 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 explain the failure of the light bulb, because, according to this fact, the light bulb should remain lit.

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 might turn out true raises the confidence in the truth of that possibility. First of all, a generation of an explanation why some future events might 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, are persevered and continued 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 explanation is absent. Brem & 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'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)

People believe in those hypotheses that, if true, would explain some event. Proponents of IBE claim, that actual explanations are those that, if true, would provide the

best explanation for some 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 (Josephson & Josephson 2003: 15; Lipton 2004: 139–140; Psillos 2002: 615). Secondly, explanatory virtues are claimed to evaluate prior probabilities and likelihoods (Lipton 2004: ch. 7; McGrew 2003; Okasha 2000; Weisberg 2009). Experiments show that people employ explanatory virtues in both of these ways.

Thagard's (1989) theory of explanatory coherence states that people accept broader, simpler and deeper explanations as better. Read & Marcus-Newhall (1993) conducted experiments to test different aspects of this theory. They found 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 & Marcus-Newhall found 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 & Marcus-Newhall found that explanations are perceived to be better when they are explained by a further explanation than when they are not. This feature of explanatory power is sometimes called a deepening of explanation (Thagard 2007) or an explication of underlying mechanism and is also considered to be an explanatory virtue operative in determining the best explanation (e.g., Psillos 2002: 615).

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 better 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 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 claims associated with 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 claimed to be one of the explanatory virtues associated with IBE. Ahn et al. (1995) examined whether people seek for information about covariance or about causal mechanism when asked to provide an explanation for some event. Experiments showed that people prefer information about underlying causal mechanism 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 associated with the activation of a particular other 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 sensation of taste, pain perception and disgust. Harris et al. concluded that the final acceptance or rejection of a statement appear 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 (2004) choice of a 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 claimed to be one of the most important things in discerning the best explanation. For Psillos (2002: 615) coherence with background knowledge is the ultimate explanatory virtue and other explanatory virtues operate only when relevant background knowledge cannot discriminate between competing explanations. For Lipton (2004: 139–140) what counts as an explanatory virtue is partially determined by background knowledge. Experiments show that background knowledge contributes to credibility of explanations and that coherence with background knowledge is a condition for a piece of information to be accepted as true

As mentioned earlier, Koslowski et al. (2008) 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 be-

come more credible when they can accommodate relevant background information. In other words, the perceived probability of explanations is a function of their coherence with the 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 show 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. found 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 found that verifying false statements activates the fronto-polar cortex and conclude that this is consistent with the claim that the processing of false statements involve 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 the background knowledge plays a decisive role when evaluating the truth of incoming information.

The often mentioned and often replicated study by Tversky & Kahneman (1982) also shows that background knowledge is relevant for perceived probability. In the Tversky & 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 & 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 some 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. Proponents of IBE only require that this mere ability to explain should be good enough.

One more experiment by Read & 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 & Marcus-Newhall found that that the perceived goodness of a set of narrow explanations was lower when a broad explanation was present than when it was absent. Hence, a presence of a better explanation lowered the perceived probability of other explanations.

The truth of the psychological hypothesis is taken for granted by some philosophers, e.g., Carruthers and Fodor. The truth of the psychological hypothesis is also assumed in the pragmatic and evolutionary justifications of IBE. However, the truth of the psychological hypothesis is often assumed without any justification of it. Therefore, the present analysis enumerated empirical evidence that fills this void.

Ontology in IBE

IBE is not deductively valid and, therefore, best explanations cannot be true in any 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. The ostensible validity of IBE is material, not formal. 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 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 world has a particular (coherent, unified and simple) ontological structure.

This biconditional means that the theory 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 would be as coherent, unified and simple as possible, but all explanations that are the most coherent, unified and simple would not be true 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 then 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 to be 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). (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. (1999: 83)

Day & 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. (1994: 282)

Thus, substantive assumptions as described in these quotes are seen as necessary, without which IBE form 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 & Kincaid even describe IBE as nothing more than the totality of these assumptions taken together.

Thagard (2007: 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 work not in every possible world. The world has to have a very specific ontological structure for IBE to be true. Hence, 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 happen, 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 favourite 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 this world should be constituted for IBE to work.

Then there are assumptions characteristic of induction in general. These state that there are regularities in the nature and that the nature is uniform, i.e. the physical possibilities and regularities in the nature should not change throughout the 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 than any kind of explanans would be impossible, because the same explanans even in the exactly the same circumstances would not be capable to account for the same explanandum.

There are two principal substantive assumptions characteristic solely to IBE. The first is the reliance on the 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 any 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 some 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, according to IBE, 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 in the world is exactly as described in the background knowledge.

Truth-conduciveness or confirmation-conduciveness of explanatory virtues is the second substantive assumption characteristic solely 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 this means that coherence, unification and simplicity have to be truth-conducive. If IBE is a form of inference that is reliable in the actual world then this 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 psychologically adequate: 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 to survive in this world. If there were even better survival-enhancing ampliative

ways of reasoning they would have replaced IBE. But, if the use of IBE helps to successfully adapt and act in this world and there is no better kind of inference, it might indicate that IBE reflects the structure of the actual world. IBE claims that explanations that satisfy the explanatory virtues the most, i.e., are the most coherent, unified, 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.

We can now specify the biconditional put forward at the beginning of the chapter.

> 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-conducive. 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 the theory of IBE true is IBE itself. For example, in the realism-scepticism 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. 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), 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 why IBE is a psychological fact. 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 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 true.

IBE, if true, seems to take 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 any 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 might succeed in is to ascertain, with the help of the natural sciences, that these assumptions do not hold.

Conclusions

1. IBE is *inter alia* a psychological theory – the experimental results mentioned in this paper show that human reasoning exhibits different features of IBE. Even though these results cannot be said to be conclusive (the studies analyzed separate aspects of IBE, not IBE itself), they do give empirical support for the psychological adequacy of IBE. Given this, the intuition endorsed by some philosophers that IBE leads to true inferences might arise not from IBE's conceptual or formal features, but rather because human psychological or psycho-physiological constitution makes us to perceive the best explanations as true. David Hume famously argued that causal reasoning is nothing more than a psychological habit. IBE too might be nothing more than a psychological habit, i.e. IBE is psychologically adequate, but it might not be true as an epistemic theory.
2. IBE is true as an epistemic theory if and only if the ontological theory it presupposes is true. More particularly, a particular ontological environment has to hold for IBE to be true: an external world has to exist; the world has to be uniform; the state of affairs as depicted in the background knowledge has to hold; the explanatory virtues that determine the goodness of an explanation have to be truth-conducive. The falsity of any of these claims would deny the truth-conduciveness of IBE. Conversely, if IBE is truth-conducive, then it cannot be the case that the world does not have the aforementioned ontological structure.

REFERENCES

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.

Barnes, E.(199). Inference to the Loveliest Explanation, *Synthese* 103(2): 251–227, doi:10.1007/BF01090049.

Brem, S. K. and Rips, L. J.(200). Explanation and Evidence in Informal Argument, *Cognitive Science* 24(4): 573–604, doi:10.1016/S0364-0213-(00)00033-1.

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.

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

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

Gabbay, D. M.; and Woods, J. 2005. *The Reach of Abduction: Insight and Trial, volume 2 of a Practical Logic of Cognitive Systems*. Amsterdam: Elsevier.

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

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.

Josephson, J. R.; and Josephson, S. G., eds. 2003. *Abductive inference: Computation, Philosophy, Technology*. Cambridge: Cambridge University Press.

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.

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

Lipton, P. 2004. *Inference to the Best Explanation*. 2nd ed. London: Routledge.

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

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.

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.

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

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.

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

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. 1958. *The Collected Papers of Charles Sanders Peirce*. Vol. 7. Cambridge, MA: Harvard University Press.

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

Psillos, S. 2002. Simply the Best: A Case for Abduction, in A. C. Kakas, and F. Sadri, eds. *Computational Logic: Logic Programming and Beyond*, vol. 2408 of Lecture Notes in Computer Science. Berlin–Heidelberg: Springer, 605–625.

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. 2009. 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*, vol. 10, 117–151. Elsevier, Amsterdam.

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.

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.

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

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

Tversky, A.; and Kahneman, D. 1982. Judgments of and by Representativeness, in *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 84–98.

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, 245–308.

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

GERIAUSIO PAAIŠKINIMO IŠVEDIMO PSICHOLOGINIS ADEKVATUMAS IR ONTOLOGINIAI ĮSIPAREIGOJIMAI

Adolfas Mackonis

S a n t r a u k a

Geriausio paaiškino išvedimas (GPI) dabartinėje analitinėje epistemologijoje ir mokslo filosofijoje yra plačiai tyrinėjama episteminė teorija, teigianti, kad hipotezės buvimas geriausiu paaiškinimu yra pakankama šios hipotezės teisingumo sąlyga. Straipsnyje teigiama, kad GPI analizuotina ne tik kaip episteminė, bet ir kaip psichologinė bei ontologinė teorija. Pirma, aptariami kognityviųjų mokslų tyrimai, kurie leidžia teigti, jog GPI teisingai aprašo žmonių samprotavimų praktiką: paaiškinimas suvokiamas kaip teisingumo požymis; teiginiai, kurie yra geresni paaiškinimai, priimami kaip labiau tikėtini; aiškinimo vertybės bei turimas žinojimas daro įtaką teiginių tikimybės vertinimui. Antra, straipsnyje aptariama, kokia turėtų būti pasaulio ontologija, kad GPI kaip episteminė teorija būtų teisinga. Realizmas apie išorinį pasaulį, gamtos vienodumas, turimo žinojimo teisingumas bei aiškinimo vertybių palankumas tiesai yra būtinos ir pakankamos sąlygos tam, kad GPI būtų teisinga.

Pagrindiniai žodžiai: geriausio paaiškinimo išvedimas, abdukcija, psichologija, ontologija, aiškinimo vertybės.

Įteikta: 2010-10-17