

# Evaluating explanatory theories

Brian Haig advocates 'inference to the best explanation', a method used by Darwin and of relevance for theory appraisal in psychology

Explanatory theories in psychology are usually evaluated by employing the hypothetico-deductive method and testing them for their predictive accuracy. The purpose of this short article is to bring an alternative approach, known as 'inference to the best explanation' (IBE), to the attention of psychologists. In doing so, it provides some methodological resources on the topic that psychologists can use to help them evaluate the explanatory worth of their theories. Despite the fact that IBE has been employed by a number of scientists (Janssen, 2003) – Darwin and Einstein among them – and extensively discussed by philosophers of science (e.g. Lipton, 2004), it has received little attention in psychology. Given that IBE is an important approach to theory appraisal, this is an omission that needs to be put right.

## Inference to the best explanation

Inference to the best explanation is based on the idea that much of what we know about the world, in both science and everyday life, is based on considerations of the explanatory worth of our beliefs. Scientists often accept theories about hidden causes of observed events because they are thought to be the best explanations of those events. This was the reasoning Darwin used in judging his theory of natural selection to be superior to the rival creationist explanation of his time (Thagard, 1978). The

methodological literature on IBE has endeavoured to unpack the idea of explanatory worth.

In contrast to the popular hypothetico-deductive method, IBE takes the relation between theory and evidence to be one of explanation, not logical entailment. This means that for IBE the ideas of explanation and evidence come together and explanatory reasoning becomes the basis for evaluating theories. Also, in contrast with the Bayesian approach to theory evaluation, advocates of IBE generally take theory evaluation to be a predominantly qualitative exercise that focuses on explanatory criteria, not a quantitative undertaking in which one assigns probabilities to theories. Given that the primary function of most theories in science is to explain empirical facts, it stands to reason that the 'explanatory goodness' of theories should count in their favour. Conversely, the explanatory failings of theories should detract from their credibility. The main point of IBE is that the theory judged to be the best explanation of the facts is taken as the theory most likely to be correct.

In what follows, I outline a well-developed method for carrying out IBE. I then briefly suggest that the hypothetico-deductive method can be combined with other evaluative considerations to provide

"for IBE the ideas of explanation and evidence come together"

a different approach to IBE. After that, I discuss the place of IBE in psychology's methods curriculum. I conclude that psychologists should include ideas about IBE in their methodological thinking, and use them to evaluate the goodness of their explanatory theories.

## The theory of explanatory coherence

The cognitive scientist Paul Thagard has developed a detailed account of IBE known as the 'theory of explanatory coherence' (TEC) (Thagard, 1992). According to TEC, IBE is essentially a matter of establishing relations of explanatory coherence between propositions within a theory. These propositions will be either claims about the evidence, or hypotheses that help explain the evidence. To infer that a theory is the best explanation is to judge it as more explanatorily coherent than its rivals. TEC is not a general theory of coherence that subsumes different forms of coherence, such as logical and probabilistic coherence. Rather, it is a theory of *explanatory* coherence where the propositions in a theory hold together because of their explanatory relations.

The determination of the explanatory coherence of a theory is made in terms of three criteria: explanatory breadth, simplicity and analogy (Thagard, 1978). The criterion of explanatory breadth is the most important criterion for choosing the best explanation. It captures the idea that a theory is more explanatorily coherent than its rivals if it explains a greater range of facts. For example, Darwin's theory of evolution explained a wide variety of facts that could not be explained by the accepted creationist explanation of the time.

The notion of simplicity is captured by the idea that preference should be given to theories that make fewer special or ad hoc assumptions. Simplicity is the most important constraint on explanatory

## references

- Capaldi, E.J. & Proctor, R.W. (2008). Are theories to be evaluated in isolation or relative to alternatives? *American Journal of Psychology*, 121, 617–641.
- Durrant, R. & Haig, B.D. (2001). How to pursue the adaptationist program in psychology. *Philosophical Psychology*, 14, 357–380.
- Eflin, J.T. & Kite, M.E. (1996). Teaching scientific reasoning through attribution. *Teaching of Psychology*, 23, 87–91.
- Freedman, E.G. (1992). Understanding scientific controversies from a computational perspective: The case of latent learning. In R.N. Giere (Ed.) *Minnesota studies in the philosophy of science*, Vol. 15 (pp.310–337). Minneapolis: University of Minnesota Press.
- Haig, B.D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *American Journal of Psychology*, 122, 219–234.
- Janssen, M. (2003). COI stories: Explanation and evidence in the history of science. *Perspectives on Science*, 10, 457–522.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Lipton, P. (2004). *Inference to the best explanation* (2nd edn). London: Routledge.
- Markus, K., Hawes, S.S. & Thasites, R. (2008). Abductive inference to psychological variables: Steiger's question and best explanations in psychopathy. *Journal of Clinical Psychology*, 64, 1069–1088.
- Proctor, R.W. & Capaldi, E.J. (2006). *Why science matters: Understanding the methods of psychological research*. London: Blackwell.

breadth; one should not sacrifice simplicity through ad hoc adjustments to a theory in order to enhance its explanatory breadth. Darwin believed that the auxiliary hypotheses he invoked to explain facts, such as the gaps in the fossil record, offered a simpler explanation than the alternative creationist account.

Finally, analogy is an important criterion because it can improve the explanation offered by a theory. Thus, the value of Darwin's theory of natural selection was enhanced by its analogical connection to the already understood process of artificial selection. Explanations are judged more coherent if they are supported by analogy to theories that scientists already find credible.

The three criteria of explanatory breadth, simplicity, and analogy are embedded in seven principles of TEC, which establish the relations of explanatory coherence. These principles are: symmetry, explanation, analogy, data priority, contradiction, competition and acceptability. As Thagard (1992) puts it, the principle of symmetry maintains that both coherence and incoherence are symmetric relations. The principle of explanation says that a hypothesis coheres with what it explains. The principle of explanation is the most important principle in determining explanatory coherence because it establishes most of the coherence relations. The principle of analogy is the same as the criterion of analogy. With the principle of data priority, the reliability of claims about observations and empirical generalisations will often be sufficient grounds for their acceptance. The principle of contradiction asserts that contradictory propositions are incoherent with each other, while the principle of competition claims that theories that explain the same evidence should normally be treated as competitors. The principle of competition allows noncontradictory theories to compete with each other. Finally, with the principle of acceptance, propositions are

accepted or rejected based on their coherence with other propositions. The overall coherence of a theory is obtained by considering the pairwise coherence relations through use of the first six principles.

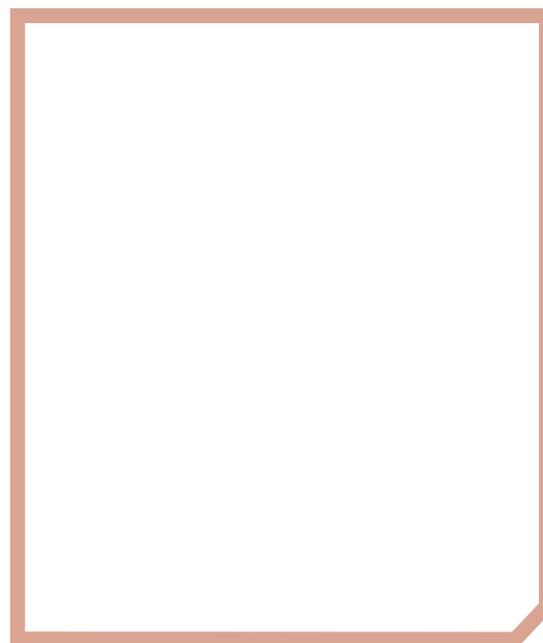
The principles of TEC combine in a computer program, ECHO (Explanatory Coherence by Harmony Optimization), to provide judgements of the explanatory coherence of competing theories. This computer program is connectionist in nature and uses parallel constraint satisfaction to accept and reject theories based on their explanatory coherence.

TEC has a number of virtues that make it an attractive theory of IBE. It satisfies the demand for justification by appealing to explanatory considerations rather than relying on predictive success; it takes theory evaluation to be a matter of comparing two or more theories in relation to the evidence; it can be readily implemented by the computer program, ECHO, while still leaving an important place for judgement by the researcher; and it effectively accounts for a number of important episodes of theory assessment in the history of science (Thagard, 1992). In short, TEC and ECHO combine in a successful formal method of IBE that enables researchers to make judgements of the best of competing explanatory theories.

It would be surprising if psychology did not contain a fair measure of competing theories that might usefully be evaluated in respect of their explanatory coherence. The use of TEC should be possible whenever two, or more, theories contain different explanatory hypotheses that purport to explain overlapping sets of empirical phenomena. In a rare use of TEC to evaluate competing theories in psychology, Freedman (1992) employed ECHO to re-examine the latent learning controversy. He contrasted the behaviourist perspective of Hull and Spence and the cognitivist approach of Tolman and his associates to cast new light on this historically important scientific controversy. This study provides an example of how one can use TEC to

reveal the complexities involved in determining whether or not reinforcement is a necessary condition for learning.

Alternatively, one might use Thagard's (1978) three criteria of explanatory breadth, simplicity and analogy in a less formal manner, as Darwin did to evaluate the worth of competing explanatory theories. For example, determining the explanatory breadth of competing theories would involve identifying the



**Archaeopteryx fossil – Darwin believed that the auxiliary hypothesis he invoked offered a simpler explanation than the alternative creationist account**

competing theories in a domain, listing all the relevant evidence statements ('observation' statements, empirical generalisations) and the explanatory hypotheses of the theories in question, and then establishing which of the two competing theories explains more evidence statements.

### **Building on the hypothetico-deductive method**

The guess-and-test strategy of the hypothetico-deductive method takes predictive accuracy as the sole criterion of theory goodness. However, it seems to be the case that in research practice the hypothetico-deductive method is sometimes combined with the use of supplementary evaluative criteria, such as simplicity, scope and fruitfulness. When this happens, and one or more of the

Rozeboom, W.W. (1999). Good science is abductive, not hypothetico-deductive. In L.L. Harlow et al. (Eds.) *What if there were no significance tests?* (pp.335–391). Hillsdale, NJ: Erlbaum.  
 Thagard, P. (1978). The best explanation: Criteria for theory choice. *Journal of Philosophy*, 75, 76–92.  
 Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.

## methods

criteria have to do with explanation, the combined approach can appropriately be regarded as a version of IBE, rather than just an augmented account of the hypothetico-deductive method. This is because the central characteristic of the hypothetico-deductive method is a relationship of logical entailment between theory and evidence, whereas with IBE the relationship is also one of explanation. The hybrid version of IBE being considered here will allow the researcher to say that a good explanatory theory will rate well on the explanatory criteria, and at the same boast a measure of predictive success. Most methodologists and scientists will agree that an explanatory theory that also makes accurate predictions will be a better theory for doing so.

Although the use of structural equation modelling in psychology often involves testing models in hypothetico-deductive fashion, it also contains a minority practice that provides an example of IBE in the sense just noted. This latter practice involves the explicit comparison of models or theories in

to evaluate the comparative merits of two- and three-factor models of psychopathy.

### IBE in the methods curriculum

For IBE to be regularly practised in psychology, the research methods curriculum will have to broaden its perspective on theory appraisal. Textbooks should present IBE as an approach to theory appraisal for psychology that is part of good scientific practice. Proctor and Capaldi's (2006) recent methods textbook *Why Science Matters* breaks new ground for psychologists in this regard. Relatedly, psychologists should be encouraged to practise IBE in their evaluation of explanatory theories, either by combining the hypothetico-deductive method with the employment of complementary evaluative criteria, as just noted, or by employing TEC. Thagard (1992) is the definitive source for a detailed explication of the theory of explanatory coherence. An introduction to using the computer programme ECHO to compute explanatory coherence can be found at Thagard's website ([tinyurl.com/ybng93y](http://tinyurl.com/ybng93y)). There, simple examples are provided that show how ECHO deals with the criteria of explanatory breadth, simplicity and analogy. Substantive examples of scientific theory choice can also be run.

Even though explicit discussions of IBE are rare in psychology, there are a few methodological papers in the psychological literature that should help researchers begin to understand different aspects of IBE. Elfin and Kite (1996) demonstrated empirically that instruction and practice in IBE improves the reasoning of psychology students in evaluating competing psychological theories. Rozeboom (1999) argued for the use in psychology of his approach to IBE, known as 'explanatory induction'. Recently Capaldi and Proctor (2008) argued, against some popular relativist trends in psychology, for the comparative appraisal of psychological theories. They recommend an approach to IBE that they call 'competing-theories abduction', where abduction has to do with explanatory reasoning. In their paper Capaldi and Proctor provide an example in experimental psychology of the use of IBE to evaluate two formal theories of attention – similarity choice theory and

signal detection theory – in respect of the relevant empirical facts. They suggest that considerations of IBE have established the fact that no other theories of attention come close to explaining the range of empirical phenomena explained by these two theories. More recently, Haig (2009) critically discussed a number of different approaches to IBE, and recommended the

"an explanatory theory that also makes accurate predictions will be a better theory for doing so"

adoption of IBE in psychological theory evaluation. As noted earlier, Markus et al. (2008) presented an understanding of structural equation modelling in terms of IBE. Finally, Durant and Haig (2001) argued that more rigorous evolutionary theories of human psychological phenomena could be achieved by employing IBE as strategy for evaluating adaptationist explanations. Although much work remains to be done to further develop the methodology of IBE, these papers should offer both the psychological researcher and the methodologist a sense of the nature of IBE and its relevance to theory appraisal.

### Conclusion

Psychology contains many competing theories that might usefully be evaluated in respect of their explanatory worth. By learning about the methodology of IBE, psychologists can position themselves to make these judgements in a more systematic way than did scientists before them. However, one should not underestimate the challenges involved in employing IBE. Apart from TEC, there are no inferential algorithms available to help researchers engage in IBE. Researchers who want to employ IBE will have to adopt more of a do-it-yourself attitude than they do in their customary use of the hypothetico-deductive method and classical statistical significance testing. Courses and workshops that focus on the use of IBE simply do not exist at present. Researchers will have to learn from the existing primary literature for themselves what the (somewhat different) approaches to IBE involve. Nevertheless, this prospect should appeal to those who want to learn about the comparative explanatory power of their theories, and use that information as a basis for accepting or rejecting them.

**I Brian Haig** is professor in the Department of Psychology at the University of Canterbury, Christchurch, New Zealand  
[brian.haig@canterbury.ac.nz](mailto:brian.haig@canterbury.ac.nz)

### Researchers may have to adopt more of a 'do-it-yourself' attitude

which an assessment of their goodness-of-fit to the empirical evidence is combined with the weighting of the fit statistics in terms of parsimony indices (Kaplan, 2000). Here goodness-of-fit provides information about the empirical adequacy of the model, whereas parsimony functions as a criterion having to do with the explanatory value of the model. Both are used in judgements of model goodness. Markus et al. (2008) recently suggested that in structural equation modelling, model fit can be combined with model parsimony, understood as explanatory power, to provide an operationalised account of IBE. They discussed the prospects of using structural equation modelling in this way

