



UvA-DARE (Digital Academic Repository)

Comparing Theories With the Ising Model of Explanatory Coherence

Maier, M.; van Dongen, N.; Borsboom, D.

DOI

[10.1037/met0000543](https://doi.org/10.1037/met0000543)

Publication date

2024

Document Version

Final published version

Published in

Psychological Methods

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Maier, M., van Dongen, N., & Borsboom, D. (2024). Comparing Theories With the Ising Model of Explanatory Coherence. *Psychological Methods*, 29(3), 519-536.
<https://doi.org/10.1037/met0000543>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Comparing Theories With the Ising Model of Explanatory Coherence

Maximilian Maier¹, Noah van Dongen^{1, 2}, and Denny Borsboom¹

¹Department of Psychology, University of Amsterdam

²Department of Philosophy and Education, University of Turin



Abstract

Theories are among the most important tools of science. Lewin (1943) already noted “There is nothing as practical as a good theory.” Although psychologists discussed problems of theory in their discipline for a long time, weak theories are still widespread in most subfields. One possible reason for this is that psychologists lack the tools to systematically assess the quality of their theories. Thagard (1989) developed a computational model for formal theory evaluation based on the concept of explanatory coherence. However, there are possible improvements to Thagard’s (1989) model and it is not available in software that psychologists typically use. Therefore, we developed a new implementation of explanatory coherence based on the Ising model. We demonstrate the capabilities of this new Ising model of Explanatory Coherence (IMEC) on several examples from psychology and other sciences. In addition, we implemented it in the R-package IMEC to assist scientists in evaluating the quality of their theories in practice.

Translational Abstract

Theories are among the most important tools of science. Lewin (1943) already noted “There is nothing as practical as a good theory.” However, although there has been a lot of discussion about improving theories, most theories in psychology are still of low quality. One reason for this is that it is difficult to assess the quality of a theory in practice. For example, when is it worthwhile to add more assumptions to explain more patterns in data? To give researchers the ability to answer questions like this, we developed a computational model for theory comparison. We also make the model available in a package for the statistical software R, so researchers can use it with ease. We hope that the availability of a tool to assess theory quality will improve the state of theory in psychology and beyond.

Keywords: theory appraisal, theory development, explanatory coherence, Ising model


Supplemental materials: <https://doi.org/10.1037/met000543.supp>


This article was published Online First March 2, 2023.

Maximilian Maier  <https://orcid.org/0000-0002-9873-6096>

Denny Borsboom and Noah van Dongen are supported by NWO Vici Grant 181.029. Code to reproduce the examples of this article is available at <https://osf.io/9wgte/>. A preprint of this article is available at <https://doi.org/10.31234/osf.io/shaef> and parts of this work have been presented at the International Workshop on Theory (Re)Construction in the Empirical Social and Behavioral Sciences. We thank Paul Thagard, Brian Haig, Freek Oude Maatman, and David Lagnado for insightful comments on a previous version of this article.

Open Access funding provided by University College London: This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0; <https://creativecommons.org/licenses/by/4.0>). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

 The data are available at <https://osf.io/9wgte/>

 The experiment materials are available at <https://osf.io/9wgte/>

Correspondence concerning this article should be addressed to Maximilian Maier, who is now at Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom. Email: maximilianmaier0401@gmail.com

Theories are among the most important tools in science. They help scientists to explain and organize knowledge (Vaidyanathan et al., 2015). They motivate the search for discoveries (Smaldino, 2019) and help us to plan interventions by predicting what would happen under hypothetical conditions (Borsboom, 2013; Fried, 2020). In addition, findings that are based on strong theory are more likely to replicate (Oberauer & Lewandowsky, 2019). Lewin (1943) put it briefly as “There is nothing as practical as a good theory” and (Thagard, 1988, p. 33) summarized it concisely as:

Scientific theories are our most important epistemic achievements. Our knowledge of individual pieces of information is little compared with theories such as general relativity and evolution by natural selection that shape our understanding of many different kinds of phenomena.

Let us be a bit more explicit. A good scientific theory, such as *evolution through natural selection*, reaches far beyond what we have observed or tested. A few elegant hypotheses explain how species evolve (see section “Comparing Creationism and Darwin’s Theory of Evolution” for details) and, in combination some ancillary hypotheses, the entire intricacy, diversity, history,

and future of biology appears to be explainable.¹ For instance, it explains why there are different species, why whales and mammals that roam the land have comparable skeletal structures, how we can domesticate plants and animals, why blue-eyed parents never have brown-eyed children, why some families have a higher than average proclivity to heart disease, why over hundreds of millions of years most species have gone extinct, why *Homo sapiens* has an appendix, why two different species cannot mate, why many but not all species reproduce through sexual intercourse, and how something as complex as binocular vision through photoreception could have evolved from unicellular animals. The theory of evolution also informs us in our development of, for instance, more resilient and nutritious crops and assists us in our fight against heritable diseases, viruses, aging, and death. The theory reaches far beyond what we are familiar with. It allows us to speculate about life on other planets that are nothing like our own—planets such as Jupiter or Venus—and reason through the possibilities for noncarbon-based or gaseous lifeforms. In short, only a relatively few interlocking hypotheses explain much of what we know about terrestrial biology and allow us to speculate about the extraterrestrial.²

However, instead of having such strong theories like *evolution through natural selection*, close to all areas of psychology are characterized by weak theories or a complete lack of theories (Fried, 2020; Oberauer & Lewandowsky, 2019). Such weak theories make imprecise predictions, often hold implicit assumptions, and are sometimes even contradictory (Borsboom, 2013; Oberauer & Lewandowsky, 2019). These problems have been discussed for decades (e.g., Gigerenzer, 1991; Meehl, 1978) and the interest in work on improving in theory is currently increasing (e.g., Borsboom et al., 2021; Fried, 2020; Guest & Martin, 2021; Haslbeck et al., 2021; Kellen, 2019; McPhetres et al., 2021; Muthukrishna & Henrich, 2019; Oberauer & Lewandowsky, 2019; Robinaugh et al., 2019; Robinaugh et al., 2021; Sanbonmatsu & Johnston, 2019; Smaldino, 2019; van Rooij, 2019; van Rooij & Baggio, 2020).

Given the long history of discussions about the need for improved theory and the comparatively little progress that has been achieved, it is important to reflect on the reasons for this situation. To assist us, let us liken theory development to navigation without a map toward a desired destination. Navigation is difficult to nearly impossible without either points of reference or instruments (e.g., compass, sextant, or GPS). Unfortunately, points of reference (empirical results) are generally far from specific and unambiguous in psychological science, for several possible reasons. First, the subject matter (e.g., intelligence) is not directly observable and fine-tuned experimental manipulation and control is difficult (or unethical). Second, to establish the field as objective science, psychologists tended to be wary of claims that are not directly scientifically testable; however, these claims are often foundational for theory development. Third, building good theories often needs to rely on simple, robust observations. However, in psychology the highest scientific credit is often attributed to those researchers that produce the most surprising and counterintuitive findings. Despite the problems of theory in psychology, thus far, most attention has been allocated to the improvement of empirical research methods (e.g., preregistration, open science, increase in statistical power, and improvement in measurement; e.g., Nosek et al., 2018; Vazire, 2018). The improvement in research methods

has shown clear benefits, though better scientific theories have not been one of them. Hence, we believe that systematic theory assessment tools (i.e., the equivalent of GPS or sextant) are required if we want to evaluate and move toward better scientific theories (Borsboom et al., 2021).

To give another example, we can also use the process of selecting the right set of predictors in a regression model as a metaphor for theory development. If we include as many predictors as data points (overfitting), we will explain all the data; however, the resulting model will not be very useful. Additionally, there is also a risk of including too few predictors (underfitting). The trade-off between overfitting and underfitting can be determined (among other approaches) using information criteria such as Akaike's information criterion (AIC). In theory development, we face a similar problem: By making our theory more complex, we can usually describe the world better; however, there is a risk of including too many assumptions. In the extreme case, a theory might have as many assumptions as phenomena that are to be explained. The resulting theory becomes merely a reformulation of empirical phenomena and will not help us discover new phenomena. However, in contrast to the case of regression, currently researchers do not have readily available tools to assess whether the inclusion of an additional assumption is justified by the increased explanatory breadth of the modified theory. One main goal of this article is to provide such a tool.³

One of the avenues we see for such tools is the assessment of the coherence between the constituent components of a theory (e.g., Haig, 2005; Poston, 2021; Thagard, 1989, 2007). A case could be made that at least part of the success of a strong theory such as *evolution by natural selection* is related to how well the theory's components interlock. Like a well-made watch where all the gears, wheels, and springs are made to minute precision to give rise to robust timekeeping without "time" being a part of any of the components. Removing or changing a single gear would radically change its behavior.

Like in a watch, the coherence of a theory's components is of vital importance. To clarify this claim, imagine one of our earlier explanations of the seasons.⁴ In ancient Greece, it was thought that winter was caused by the goddess Demeter taking the warmth from the earth, because of her grief for her Daughter Persephone spending 6 months with her husband, the god Hades, in the underworld. The problem with this theory is not that it fails to explain the lack of seasons on the equator and reversal of the seasons on the other hemisphere. The problem with this theory is that all its components can be changed without affecting the results. For instance, the gods could be completely different, the reasons for taking the warmth away could be different, or even more

¹For instance, hypotheses concerning (historical) environment and inorganic chemistry.

²We realize that certain phenomena such as human cognition as a product of biological evolution is yet to be explained, though we do not see any reason why this should not be possible in the future. Fortunately for us, the veracity of this claim has no bearing on the adequacy of IMEC.

³This metaphor is more fitting for researchers that take an instrumentalist view of theories (i.e., construe theories as a means to discover new phenomena rather than intrinsically valuable).

⁴The example that follows is a paraphrasing of the example given by David Deutsch in his inspiring book *The Beginning of Infinity* (pages 19 through 25).

fundamental changes, such as bringing warmth instead of taking it away or removing Persephone and Hades from the story, would not have to change anything in the result of seasons.⁵

On the other hand, the conventional theory of seasons does not have any components that can be changed without major repercussions. As we know it, (a) heat radiates from the sun; (b) the earth is a sphere that periodically rotates around the sun and around its axis; (c) the axis of the earth is tilted and through this axis tilt, the radiation of the sun hits the earth at an oblique angle; and (d) the atmosphere of the earth shields its surface from a certain amount of the radiation depending on its density. When these components are combined, winter on the northern hemisphere is when, during earth's revolution around the sun, the northern hemisphere is tilted away from the sun, the sun's radiation travels a longer distance through the earth's atmosphere and is heating the surface in a smaller amount. This *axis-tilt theory of seasons* explains why there are annual seasons, why there is a half-year difference between the northern and southern hemispheres, and why there are no seasons on the equator. Removing or changing one of the components of the theory would have drastic consequences; the annual seasons as we know them would not arise. In brief, strong scientific theories show a high degree of *explanatory coherence* (Thagard, 1989) between its constituent components, which could be considered an informative quality to assess. In addition, the axis-tilt theory has more explanatory breadth, because it also explains (potential) seasons on other planets in other solar systems and why the seasons in the northern and southern hemisphere of any planet with an axis-tilt are out of phase. The value of such explanatory breadth is also captured by explanatory coherence (see the section on Explanatory Breadth).

Building on Thagard's foundational work, in this article we introduce a systematic approach to the evaluation of scientific theory: the Ising Model of Explanatory Coherence (IMEC).⁶ This tool serves three purposes. First, it can help scientists to assess the qualities of novel theories that they develop, and as such play a constructive role in theory formation. Second, it allows for the systematic evaluation of existing theories in a field, so that it can support researchers in theory choice. Third, the tool can assist researchers in explicating and pinpointing problems in existing theory and as such support improvement of psychological theory.

The structure of this article is as follows. First, we explain the difference between data, phenomena, and theory, because this is a key distinction in our approach. This is followed by an explication of the theory of explanatory coherence (TEC), a way to assess the quality of a theory based on establishing principles (Thagard, 1992). On this foundation, we develop a new implementation of TEC: the Ising model of Explanatory Coherence. In the second to last section, we exhibit the use of IMEC and how it expresses desirable theoretical qualities using simple examples as well as cases from the history of science to validate and illustrate the model. Specifically, we first demonstrate that uncontroversial theoretical qualities (e.g., simplicity) are embodied by IMEC. Second, we compare two different theories explaining the *positive manifold of intelligence* either by mutualism or a common (latent) cause, to show how IMEC can be used to compare different psychological theories and identify critical experiments. Third, we compare *intelligent design* (formerly known as *creationism*) to *evolution by natural selection* to illustrate how the explanatory coherence of a theory is influenced by alternative explanations.⁷ The article is

concluded with a discussion of the limitations and suggestions for further development and research.

Theory, Phenomena, and Data

For our method, we rely on a distinction between data, phenomena, and theory (Bogen & Woodward, 1988; Haig, 2005; van Rooij & Baggio, 2020; J. F. Woodward, 2011; Figure 1). In this distinction, *data* are structured representations of observations that can provide evidence for phenomena. In psychology, typical data are reaction times, error rates, responses to multiple-choice questions, and scores on Likert or visual analog scales. *Phenomena* are general and stable features of nature, which scientists seek to explain (Bogen & Woodward, 1988; J. Woodward, 1989; J. F. Woodward, 2011). We tend to identify phenomena as general patterns in data—structures that are observed across data sets—that, in psychology, are often called “effects.” A prime example of a phenomenon is the *Stroop-effect* (e.g., MacLeod, 1991). The Stroop-effect is the general pattern that people take longer to classify the color of a word when the word spells an incongruent color. To evaluate the relationship between data and phenomena, researchers usually use statistics. For example, a significant interaction between condition and reaction time (RT)/error rates in a within-subject analysis of variance (ANOVA) could indicate evidence for the Stroop-effect. *Theories*, on the other hand, do not directly relate to the data but explain existing phenomena and motivate the search for new phenomena. For example, *automaticity*—people need to suppress the automated reading of a word—is a theory that explains the phenomenon that reaction times are longer in incongruent trials. A theory can be considered a set of hypotheses that collectively explain the phenomena.⁸ In the case of automaticity, one of its hypotheses would be that suppressing an automated processes requires attention and time, which explains the longer RT for incongruent tasks.

The process of moving between theories, phenomena, and data can be seen as an inference cycle (see Figure 1). From theories, we can deduce phenomena that nature should exhibit if our theory is correct. These phenomena in turn give use predictions about what patterns we should observe in data from observation and experimentation. In turn, from particular data sets that share common patterns, we can generalize to phenomena. From phenomena, we can conjecture a theory that would explain these phenomena (i.e., abduction; Haig, 2005; Peirce, 1958; Thagard, 1992; Thagard & Nowak, 1988). As scientists, we would like to evaluate how well data, phenomena, and theory are connected. Applied statistics offers an abundance of techniques to evaluate the extent to which phenomena are supported by the data. However, few methods exist for appraising how well phenomena support or follow from theories.

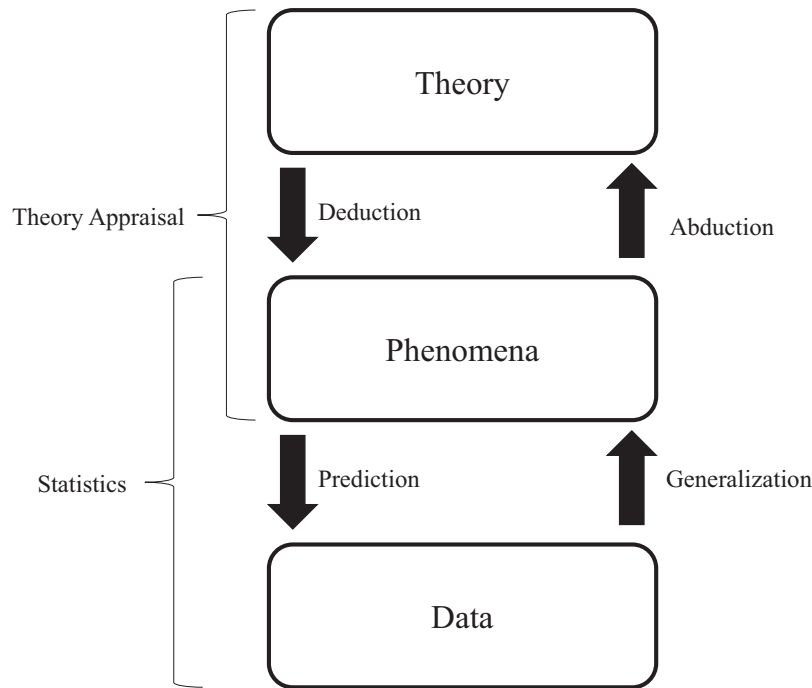
⁵ Please note that this example can easily be replaced by *intelligent design*, the *élan vital* theory, *ego depletion*, or many others that were considered scientific at some point in history.

⁶ To facilitate application in practice, we implemented our new method in the R package. A tutorial for fitting IMEC in R can be found at <https://osf.io/j4uvw/>.

⁷ More example applications can also be found in the Appendix.

⁸ Note that these are typically substantive explanatory hypotheses like “organic beings are subject to natural selection” and are not to be confused with typical statistical hypotheses, which merely involve the value of a parameter.

Figure 1
Relationship Between Theory, Data, and Phenomena



The Theory of Explanatory Coherence

Thagard (1989, 1992) proposed such a theory appraisal method; a systematic way to compare different theories based on *explanatory coherence*. Explanatory coherence is a concept devised to capture several considerations of scientific reasoning when appraising theories. It is geared toward capturing a theory's consistency with established theories and the breadth of what a theory is said to explain (i.e., number of phenomena), which is in turn constrained by its simplicity. This embodies crucial aspects that correct theories agree, that theories should not be too simple and fail to explain anything and that theories should be penalized for the addition of ad-hoc assumptions to explain more phenomena (Thagard, 1989, 1992).

For explanatory coherence, a theory consists of a set of explanatory hypotheses that explain a set of phenomena that are supported by empirical research. Within this framework, the evaluation of a hypothesis is based on the number of explanatory relations to other hypotheses or claims about phenomena. For example, if a hypothesis explains multiple phenomena, it should be preferred over a hypothesis that explains only one phenomenon. Let us illustrate this using the Stroop task and attention example from above. The fact that suppressing an automated processes requires attention, not only explains longer reaction times on incongruent trials, but also higher error rates as well as impaired performance under high cognitive load (MacLeod, 1991). This makes it preferable over a theory that would need to postulate different hypotheses to explain each phenomenon separately. To capture intuitive principles like these, Thagard (1989; revised in Thagard, 1992) postulated seven principles that give rise to the explanatory coherence relationships. From these explanatory coherence relations the theory appraisal considerations "follow as a matter of course." These principles

form the (meta) *theory of explanatory coherence* (TEC).⁹ We quote these seven relations from Thagard (2000, p. 43):

1. **Symmetry.** Explanatory coherence is a symmetric relation, unlike, say, conditional probability. That is, two propositions *p* and *q* cohere with each other equally.
2. **Explanation.** (a) A hypothesis coheres with what it explains, which can either be evidence or another hypothesis. (b) Hypotheses that together explain some other proposition cohere with each other. (c) The more hypotheses it takes to explain something, the lower the degree of coherence.
3. **Analogy.** Similar hypotheses that explain similar pieces of evidence cohere.
4. **Data Priority.** Propositions that describe the results of observations (usually claims about phenomena) have a degree of acceptability on their own.
5. **Contradiction.** Contradictory propositions are incoherent with each other.
6. **Competition.** If *p* and *q* both explain a proposition, and if *p* and *q* are not explanatorily connected, then *p* and *q* are incoherent with each other (*p* and *q* are explanatorily connected if one explains the other or if together they explain something).

⁹ These seven principles can also be seen as meta-hypotheses, which can in principle be tested and evaluated like regular hypotheses.

7. **Acceptability.** The acceptability of a proposition in a system of propositions depends on its coherence with them.

In these principles, a proposition is either an explanatory hypothesis of a theory, or a phenomenon, which is to be explained by a theory. The principle of explanation is the most important principle because it establishes most of the explanatory coherence relationships. Acceptance can be considered the dependent variable of TEC, which results from the explanations established by the other six principles. See Thagard (1992) for a detailed disquisition of all seven principles and their origin.

Thagard (1989) implemented these principles in a connectionist model called Explanatory Coherence by Harmony Optimization (ECHO).¹⁰ Connectionist models describe a network with excitatory and inhibitory links between nodes. The different nodes can have varying degrees of activation. Whenever a given node is activated it activates nodes with which it is connected through excitatory links and deactivates nodes with which it is connected through inhibitory links. The degree to which the connected node is activated or deactivated depends on the strength of the excitatory or inhibitory link. In ECHO, two (or more) theories are compared in their explanation of phenomena. The network of these phenomena and hypotheses of the theories is run in several iterations in which nodes update each other according to the connecting links, usually until it settles in a stable state where one theory has a higher activation than the other(s). A detailed analysis of ECHO is beyond the scope of this article (see Thagard, 1989, 1992).

ECHO has proven its value in the appraisal of theories, yet it has two key limitations.¹¹ The first limitation of ECHO is that, because of the connectionist architecture, it is incapable of appraising individual theories. It can only be used to compare two or more theories. This is because, when only one theory in relation to the phenomena is being evaluated in ECHO, only excitatory links will be established because contradictions (inhibitory links) usually only occur between two competing theories. Therefore, the activation of nodes will increase indefinitely when running the network for several rounds. The system does not settle to a stable state that expresses the explanatory coherence. ECHO's second limitation is that it is not available in any software psychologists typically use, which greatly reduces its usability for theory improvement in psychological science. These limitations prompted us to develop a novel implementation of TEC.

The Ising Model of Explanatory Coherence

We developed our instrument in the common programming language R and implemented TEC using the Ising model (Ising, 1925). This has resulted in an R package IMEC that researchers can easily use to appraise the quality of their theories (Maier, 2021).

Revising TEC

For the implementation of TEC in IMEC, we exclude several of TEC's principles. Specifically, we excluded Principles 2b (propositions that explain evidence together cohere), 3 (analogy), and 6 (competition). Thagard (1989) based his Principle 2b (part of Explanation) "For each P_i and P_j in $P_1 \dots P_m$, P_i and P_j cohere" on the Duhem-Quine thesis (Duhem, 1954) that explanation or prediction requires a bundle of mutually dependent cohypotheses and

assumptions. However, Thagard (1989, p. 437) already notes problems with this principle. He states "any scientist who maintained at a conference that the theory of general relativity and today's baseball scores together explain the motion of planets would be laughed off the podium. Principle 2 is intended to apply to explanations and hypotheses actually proposed by scientists." There are many more cases where we can find problems with this principle. For example, if a depression is caused by one's physiological predisposition as well as one's life partner dying in a car crash, would this lead us to conclude that physiological predisposition for depression and life partners dying in car crashes cohere? It seems more that two *unrelated* causes together explain the depression. Therefore, we decided not to model Principle 2b.¹²

We also decided against modeling Principle 3, Analogy. While analogy often seems to be useful in generating new discoveries, it is not established as a criterion for the quality of a theory. In practice, it is also often hard to differentiate what kinds of explanations are actual analogies and what kinds are illustrative metaphors. Noteworthy for psychology, some theories that were based on analogies or metaphors turned out to be hardly replicable (e.g., Carter & McCullough, 2014). A case could be made that actual analogies, as differentiated from metaphors, increase the approximate truth of a theory. An example is the analogy between artificial selection of traits in the domestication of herd animals and the development of traits in animals through natural selection. However, without giving a detailed explication of what makes something sufficiently count as an analogy, scientists can easily "boost" their theory's explanatory coherence by tagging on metaphors and selling them as analogies. Such demarcation is beyond the scope of the article; thus, we (currently) did not model analogy in IMEC.

Finally, we excluded Principle 6, the principle of competition "If P and Q both explain a proposition P_i , and if P and Q are not explanatorily connected, then P and Q incohere." This principle seems to be a requirement of the connectionist architecture of ECHO. Namely, if there are too many excitatory links compared with inhibitory links in a network, then the activation of nodes increases with each iteration and the network does not settle in a stable state. Hence, Thagard (1992) seems to have added the principle of competition to create more inhibitory links. In contrast to ECHO, IMEC does perform well with excitatory links only. Therefore, we did not incorporate the principle of competition by default. However, researchers who want to incorporate the principle of competition can do so by adding inhibitory links between propositions of different theories that explain the same phenomenon. This absence of an inhibitory link requirement allows researchers to evaluate individual theories. In addition, some have advocated to use IMEC analogous to models of probabilistic reasoning such as Bayesian networks (Dahlman & Mackor, 2019). In this case the principle of competition can also be useful to model "explaining

¹⁰ To acknowledge Gilbert Harman, who conceptualized inference to the best explanation, Thagard uses the word Harmony rather than the more common word harmony.

¹¹ For instance, it proved its mettle in evaluating key theories from the history of science, such as comparing different theories about why the dinosaurs became extinct or different theories to explain continental drift (Nowak & Thagard, 1992; Ranney & Thagard, 1988; Thagard, 1991, 2005).

¹² Researchers who want to incorporate this principle when using IMEC can still do so by manually adding explanatory links between hypotheses that together explain phenomena.

away.” If there are two independent explanations for a phenomenon and we learn that one of them is true, the other one will become less likely to be true. Therefore, in situations where increased plausibility of one explanation necessarily leads to decreased plausibility of another, we recommend adding a negative link between the relevant explanatory hypotheses. However, in practice, it is often difficult to know whether two causes are independent and hence whether the situation involves explaining away, which is why we did not model the principle of competition by default.

Requirements for a Model That Can Implement the Modified TEC

We are left with five out of seven of TEC’s principles. These principles are to be our meta explanatory hypotheses from which the qualities of good theories should follow as a matter of course. We summarize these principles as the following requirements of a model that incorporates TEC:

1. The model needs symmetric connections between hypotheses and phenomena to incorporate Principle 1 (symmetry).
2. It must be possible to establish positive as well as negative connections between phenomena and hypotheses to represent Principle 2 (explanation).
3. It needs some external force that gives the phenomena some degree of acceptability on their own to incorporate Principle 4 (data priority).
4. It must be possible to establish negative connection to represent Principle 5 (contradiction).
5. It requires a measure for the activation of phenomena and hypotheses to denote Principle 7 (acceptability).

Incorporating TEC in the Ising Model

One model that has these four properties is the Ising model (Ising, 1925). We developed a new computational algorithm to evaluate explanatory coherence based on the Ising model that we term the IMEC. To understand the proposed method, it is necessary to outline the basic properties of the Ising model.

The Ising model was originally developed in statistical mechanics to describe the polarization of ferromagnetic materials (Ising, 1925). However, its properties are useful to model a variety of systems and the model has been applied widely, for example, to model psychological constructs and opinion dynamics (Dalege et al., 2016; Epskamp et al., 2016; van der Maas et al., 2020). The Ising model describes a binary network in which each node can have states 1 and -1 . This can be interpreted as the node being either “on” or “off.” For instance, a person either has a symptom or she does not. The probability of a node being in a certain state is influenced by three parameters (Epskamp et al., 2016).

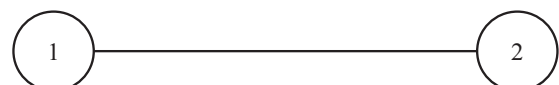
1. The threshold parameter τ represents the tendency of a node to be in a given state. The higher (lower) τ , the more the node prefers to be in state 1 (-1). For example, in

Figure 2, if we add a threshold of one to node one, this would increase the probability of this node to be in state one. The probability would be 0.881 and, thus, higher than the 0.5 probability of being in state one that we would find without threshold.

2. The network parameter or edge weight ω denotes the pairwise interaction between two nodes. A positive ω between two nodes indicates that they prefer to be in the same state, while a negative ω between two nodes indicates that they prefer to be in different states. The larger the absolute value of ω the stronger this tendency. In Figure 2 adding the threshold on node one does not only increase the probability of node one to be in state one but also of node two. The reason is that this node prefers to be in the same state as node one because of their positive connection. Thus, after adding the positive threshold to node one, the probability of node two to be in state one is also increased to 0.790.
3. The inverse temperature β determines the entropy of the model and is only of minor importance for our implementation. If β equals zero, the probability of each state is equal, whereas for very high values of β , the network can often be in only one state. Note that β is a scalar on ω and τ and can be modeled using these parameters. In other words, if we divide β by some value and multiply all ω and τ parameters with the same value, we obtain the same probability (Epskamp et al., 2016). For instance, a node with $\tau = 1$ and $\beta = 1$ has the same probability of being in a given state as a node with $\tau = 2$ and $\beta = 0.5$). Therefore, we do not need to model β directly.

We can think of the different nodes as propositions in terms of TEC. In other words, each node represents either a phenomenon or a hypothesis in a theory. The relationship between phenomena and hypothesis and their acceptability can then be determined using the principles of explanatory coherence. The pairwise interactions are by definition symmetric and, therefore, satisfy Principle 1 (symmetry). Principles 2 (explanation) and 5 (contradiction) are incorporated through the edge weights ω that allow us to establish positive and negative connections between nodes. In addition, to incorporate Principle 2 (c) that the degree of coherence is inversely proportional to the number of explanatory hypotheses, we divide the edge weights by the number of hypotheses needed for an explanation (e.g., if two hypotheses are needed to explain a phenomenon the edge weight is half compared with a situation where only one hypothesis is needed to explain a phenomenon). Lower edge weights imply that the hypotheses become less activated if the phenomenon is in an activated state (i.e., is represented as true). For instance, the hypotheses that organic beings have evolved and

Figure 2
A Simple Example Case to Illustrate The Ising Model



that organic beings undergo natural selection together explain that the embryos of different species are similar; thus, they would each receive half of the activation generated from the phenomenon.

Third, by incorporating a threshold τ parameter for the phenomena, we can model the impact of evidence, or data priority (Principle 4). Thus, the threshold parameter acts to represent an external field of empirical evidence that impacts nodes representing phenomena in a way that is independent of the theory: this leads empirical phenomena to have plausibility of their own. In the example above, the stronger the evidence for the similarity of embryos of different species, the stronger would be the threshold for this phenomenon. This would also imply that the hypotheses connected to this phenomenon (organic beings have evolved, organic beings are subject to natural selection) are also supported more strongly. In cases where there is no specific knowledge regarding the strength of evidence and strength of explanation, we use default settings inspired by Thagard (1989), which are thresholds of two on the phenomena, a weight of one for explanation, and a weight of minus four for contradiction.

Finally, we need a measure for the explanatory coherence, which is manifest in Principle 7 (acceptability). This explanatory coherence is the dependent measure of IMEC. Theories with higher explanatory coherence are preferable over theories with lower explanatory coherence. In the Ising model, we can consider the explanatory coherence of a state as the probability that a state occurs. States that are more likely to occur in the Ising model have higher explanatory coherence than states that are less likely to occur.

The probability of a state (i.e., a given configuration of *all* nodes) in the Ising model can be calculated with the following equation:

$$Pr(X = x) = \frac{\exp(-\beta H(x))}{Z}. \quad (1)$$

Here, Z denotes the sum of $\exp(-\beta H(x))$ over all possible states. $H(x)$ simply denotes the Hamiltonian function. This is a function of the form

$$H(x) = -\sum_i \tau_i x_i - \sum_{i,j} \omega_{ij} x_i x_j, \quad (2)$$

where τ_i denotes the thresholds on the node and ω_{ij} the connection to other nodes for a state x .

The explanatory coherence (defined by Principle 7 or acceptability) of a hypothesis can be represented by the probability of a given node to be in State 1. Therefore, we sum $Pr(X = x)$ over all states X where a given node X_k is in state one. The explanatory coherence can be denoted as.

$$ec(X_k) = Pr(X_k = 1) = \sum_{x \in (x_k=1)} \frac{\exp(-\beta H(x))}{Z}. \quad (3)$$

In other words, we calculate the probability of all given configurations in which a hypothesis is true.¹³ If we look at a set of hypotheses, rather than a single one, the explanatory coherence is

simply denoted as the *average* over the coherence of the individual hypotheses. Although we use the probability of a given state in the Ising model as the foundation for our method, we want to emphasize here, that the resulting number **cannot** directly be interpreted as the probability of a theory being true. Different considerations, such as the role of contradictions, and the fact that we average over individual nodes, when calculating the coherence of a theory complicate this matter. Therefore, the explanatory coherence values should simply regarded as numbers between zero and one, and not as probabilities for the truth of a theory. In the next section we give some aid for the interpretation of these numbers.

Rule of Thumb for Interpreting Explanatory Coherence

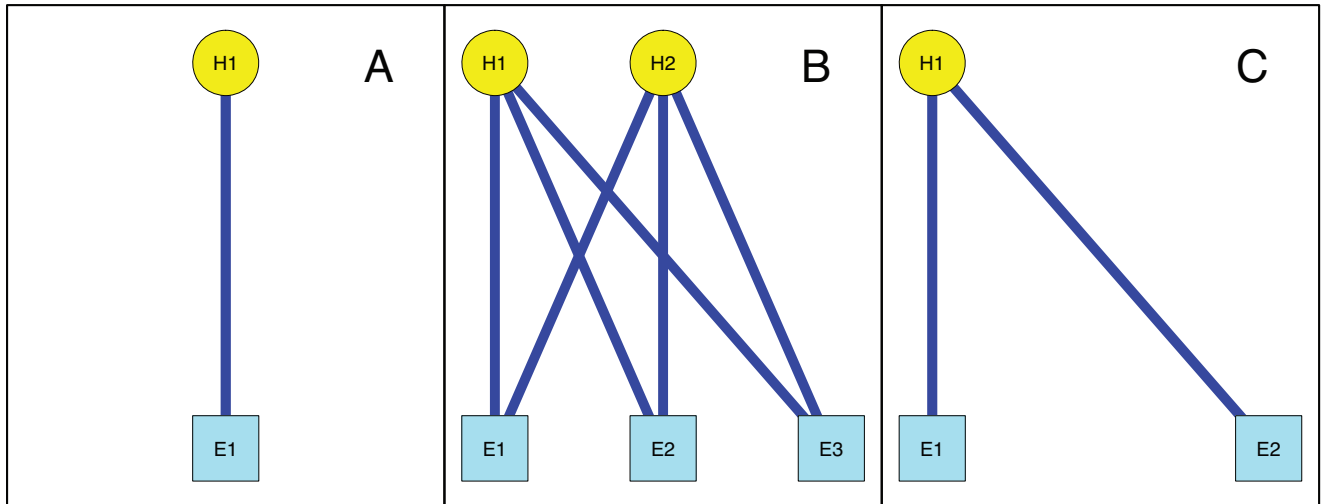
The higher this explanatory coherence is, the better, but how high is enough to actually accept a theory? The question of how to interpret the values for the explanatory coherence of theories requires careful consideration. We do not want to establish a default benchmark. Plenty of problems have been ascribed to blindly following the p -value threshold of .05 (e.g., Lakens et al., 2018; McShane et al., 2019) without paying the necessary attention to context (e.g., sample size and effect size). In addition, what determines a good theory depends on many contextual factors, such as the typical quality of theories in a field of research. Therefore, we cannot offer default benchmarks to interpret explanatory coherence. At the moment, we can offer two approaches to construct rules of thumb, one for single theory evaluation and one for theory comparison. In the single theory case, we can derive reasonable values based on the hypotheses: phenomena ratio of typical theories in the field. Although these considerations are highly specific to the chosen research field and should, therefore, always be evaluated in context Figure 3 may serve as a simple example to illustrate a procedure of deriving explanatory coherence values. This indicates that an explanatory coherence of .867 can be considered low, an explanatory coherence .950 can be considered medium, and an explanatory coherence of .977 can be considered high. However, we want to reiterate that these benchmarks are only meaningful in the single theory case. In the theory comparison case, contradictions between theories complicate the picture and these rules of thumb should not be used. In the case of theory comparison, researchers should favor the theory with the highest explanatory coherence.

IMEC's Representation of Different Criteria for Theory Evaluation

Next, let us show on four small examples inspired by Thagard (1989) how IMEC smoothly integrates considerations that scientists use for theory evaluation in practice. These are explanatory breadth, refutation, simplicity, and deactivating evidence (e.g., Meehl, 1992, 2002). In the next examples we run IMEC with default settings inspired by Thagard (1989) and validated by re-analyzing his work. This is a threshold (evidence) of two on the phenomena, positive weights of one for the explanations, and weights of minus four for contradictions. While this setting has been

¹³ Note that this explanatory coherence can not only be calculated for hypotheses but also for the claims about phenomena themselves, although they often have higher explanatory coherence due to data priority.

Figure 3
Rule of Thumb for Interpreting Explanatory Coherence in the Single Theory Case



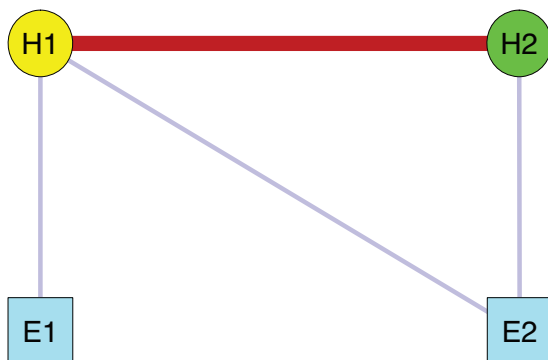
Note. (A) Low explanatory coherence (0.867); (B) Medium explanatory coherence (0.950); (C) high explanatory coherence (0.977). See the online article for the color version of this figure.

shown suitable in comparing a wide range of cases, there is nothing special about these parameter values. In practice, we encourage researchers to deviate from the default if, for example, one phenomenon has stronger empirical support than others.

Explanatory Breadth

Explanatory breadth, the number of phenomena that a theory can explain, is a criterion scientists commonly use to compare theories. If a theory explains more, it is preferred over a theory that explains less. Therefore, given the same complexity, IMEC should prefer the proposition that explains more phenomena. This setting is illustrated in Figure 4, where H1 gives a broader explanation of observed phenomena

Figure 4
Explanatory Breadth



Note. Blue (thin) lines indicate positive connections with weight 1, red (thick) lines indicate negative connections with weight 4 and thresholds on the phenomena are set to 2. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

and should, thus, be preferable. Let us consider how IMEC evaluates this case. Comparing the two theories with IMEC based on Equation 3 indicates an explanatory coherence of .867 for H1 and an explanatory coherence of .134 for H2. In other words, IMEC seems to correctly prefer theories of higher explanatory breadth.

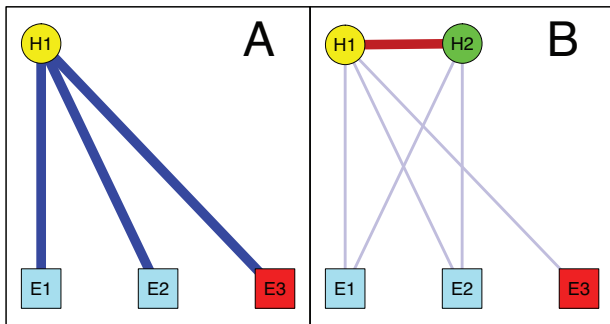
Refutation

Falsification—the abandoning of refuted theories—is sometimes considered the defining criterion of science (Popper, 1959). However, in practice scientists do not directly discard an otherwise successful theory after one failed prediction and it is probably also wise not to do so (Lakatos, 1978). The refutation of a theory might often depend more on whether better, alternative explanations become available. Figure 5A shows a setting where a proposition H1 explains two pieces of evidence E1 and E2 but also predicts a phenomenon E3 for which contradicting evidence exists. We use the same thresholds and edge weights as before but assign a threshold of -1 to E3. In this setting, IMEC assigns an explanatory coherence of .867 to this hypothesis. In other words, the hypothesis is not directly refuted because of one incorrect prediction. Figure 5B shows the same setting, however, now an alternative explanation H2 has become available which does not predict E3 and contradicts the old explanation. Now, the explanatory coherence of H1 drops to .135 and H2's explanatory coherence is .867. This shows that the explanatory coherence of a proposition usually depends also on the availability of alternative explanations.

Simplicity

Because of Principle 2c, the connection between a phenomenon and its explaining proposition is inversely proportional to the number of explanations. Therefore, if more propositions are needed to explain a phenomenon, this should result in lower explanatory coherence,

Figure 5
Refutation



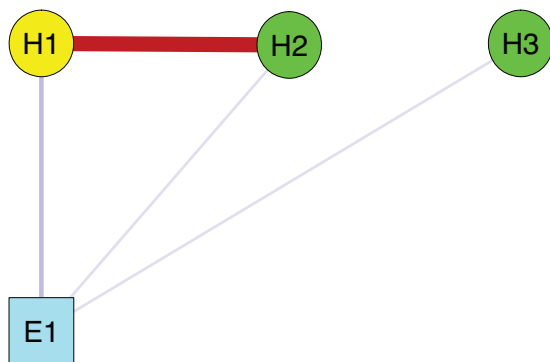
Note. Blue (thin) lines indicate positive connections with weight 1, red (thick) lines indicate negative connections with weight 4. Thresholds on the phenomena are set to 2 for E1 and E2 and to -2 for E3. Note that the blue weights are equal in both figures and are scaled down in figure B only in comparison with the contradiction of weight 4. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

reflecting the principle of simplicity. Figure 6 shows two theories with either one proposition (yellow) or two propositions (green). Because the green theory needs two propositions to explain E1 each of them only has a connection of $1/2 = .5$ to this phenomenon.¹⁴ All other weights and thresholds are one or minus one as before. With this setting, IMEC indicates that the simpler theory's explanatory coherence is .722, while the explanatory coherence of the two proposition theory is only .500. In other words, IMEC captures simplicity as a criterion for theory appraisal.

Deactivating Evidence

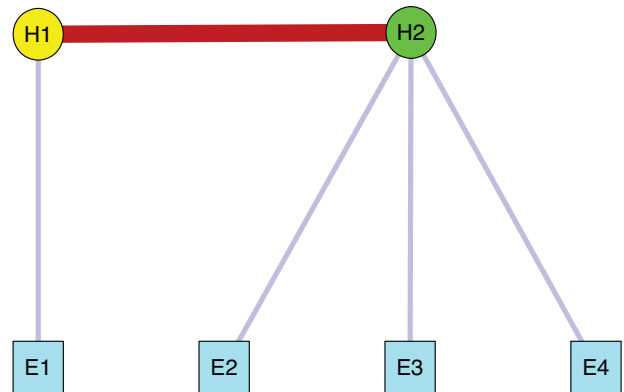
Because of the principle of data priority, phenomena that are supported by evidence have some explanatory coherence on their own as modeled by assigning them positive thresholds. However, if a

Figure 6
Simplicity



Note. Blue (thin) lines indicate positive connections, red (thick) lines indicate negative connections. The connections from H2 and H3 to E1 are only 0.5 in this setting because they jointly explain this phenomenon. Thresholds are set to two. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

Figure 7
Deactivating Evidence



Note. Blue (thin) lines indicate positive connections with weight 1, red (thick) lines indicate negative connections with weight -4 and thresholds on the phenomena are set to 2. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

phenomenon contradicts a variety of well-supported propositions, researchers often disbelieve it or downplay this phenomenon. Let us consider the example in Figure 7. We can see that the phenomenon E1 is connected to the inferior H1 that contradicts a better supported hypothesis (H2). All edge weights are set to 1 and thresholds are set to 2 or -4 (red weights). In this setting, the explanatory coherence for phenomenon E1 is only .884 and smaller than for the other phenomena (.995, .995, and .995). In other words, the phenomenon was downplayed because it was only predicted by one very implausible proposition H1.¹⁵

Example Applications of IMEC

Positive Manifold of Intelligence by Mutualism

The next paragraph shows the usefulness of IMEC as a tool to compare psychological theories by comparing the explanatory coherence of two theories of intelligence, the *g*-factor explanation (e.g., Spearman, 1904) and mutualism explanation (van der Maas et al., 2006). We also show that IMEC allows to evaluate the robustness of the theory to weak evidence or weak explanations and to identify critical experiments by thinking through counterfactuals.

Van der Maas et al. (2006) propose an alternative model of intelligence that explains the positive manifold (the positive correlation of different components of intelligence) not by commonly used latent variable explanations (*g*-factor) but by mutualism. Development by mutualism means that the positive correlation between different aspects of intelligence (e.g., verbal reasoning, logical-mathematical thinking) occurs solely due to positive interactions between several distinct cognitive processes during development. This idea is inspired by ecology, where for example, the

¹⁴ In practice IMEC will do this kind of scaling automatically.

¹⁵ Note, if the researcher finds this property of downplaying evidence undesirable, it can easily be prevented by assigning high thresholds to the phenomena.

correlation between different aspects of water quality in a lake (e.g., vegetation, water quality) is not explained by a “lake-factor” but by the positive interactions between different aspects of water quality (e.g., Scheffer, 1997; Scheffer et al., 1993).

Based on this model, van der Maas et al. (2006) explain a variety of phenomena in intelligence research, some of which latent variable models have struggled to explain for a long time. However, they also introduce some new assumptions (explanatory hypotheses that support the main hypothesis); therefore, it is interesting to investigate whether their new theory has more explanatory coherence than a latent variable explanation.

Table 1 shows different phenomena related to intelligence and the propositions of the mutualism and the *g*-factor explanations. The explanatory relations can be seen in Figure 8 with edge weights and thresholds based on the default explained above. Looking at the figure shows that it appears difficult to compare these two theories intuitively. The mutualism theory explains more phenomena than the *g*-factor theory. However, it is also more complex. Hence, an instrument like IMEC that can assess the trade-off between these different epistemic values seems required for effective theory comparison.

Implementing these explanatory relations in IMEC indicates that the mutualism explanation is superior with an explanatory coherence of .788, whereas the explanatory coherence of the latent variable explanation is only .504; in other words, the mutualism hypothesis seems preferable. This exemplifies how IMEC can smoothly compare theories in contexts where it is intuitively hard to decide between theories.

However, van der Maas et al. (2006) state regarding E7 (differentiation effects) that this phenomenon has not been replicated consistently. Differentiation effects imply that the *g*-factor is not uniform in the population. In particular it has been suggested, that the positive manifold declines with age (e.g., Tideman & Gustafsson, 2004) and that the positive manifold is stronger in lower IQ groups (e.g., Deary et al., 1996). However, both of these manifestations could sometimes not be replicated (e.g., Facon, 2004). In other words, due to the weak evidence for this phenomenon, we

Table 1
Mutualism Versus Latent Variable Explanation

Phenomena
E1: There exists a positive manifold between test scores.
E2: Intelligence is hard to predict from early childhood performance.
E3: Intelligence can be described by a hierarchical factor structure.
E4: General intelligence is highly heritable.
E5: Jensen effect - The larger the <i>g</i> -loading of cognitive tests the larger the heritability.
E6: Flynn effect - IQ increased over time.
E7: Differentiation effect - The positive manifold is not uniform in the population.
E8: No biological variable could be identified as the unitary cause of <i>g</i>
Mutualism hypotheses
HM1: Each cognitive process supports the development of other processes.
HM2: Growth can be described by a logistic model.
HM3: There are small genetic intercorrelations of resources.
HM4: Intelligence and environment cause each other reciprocally.
Latent variable hypotheses
HL1: Intelligence is caused by a common latent variable.
HL2: Intelligence and environment cause each other reciprocally.

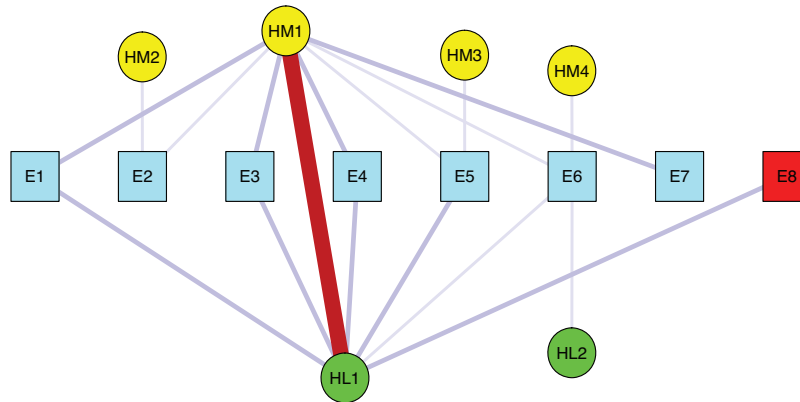
Note. Extracted from van der Maas et al. (2006).

should consider stepping away from IMEC’s default settings and lowering the threshold on E7. In addition, van der Maas et al. (2006) state that the mutualism model allows for both differentiation and integration; in other words, the model makes a very general prediction with regards to differentiation effects that can easily be confirmed. Therefore, let us consider what happens if we reduce both the evidence for E7 and the weight between HM1 and E7 from 1 to .5. In other words, to account for the problematic state of this phenomenon and the vague prediction of the mutualism model, we give only half the evidence to the phenomenon E7 and we weaken the connection to show that the phenomenon is implied weaker by the theory than other phenomena. Computing IMEC with these settings results in an explanatory coherence of .779 for the mutualism model and an explanatory coherence of .516 for the latent variable model; a change of $-.009$ and $.008$, respectively. In other words, the superiority of the mutualism model seems to be robust to the problems with differentiation effects. This example shows how we can incorporate weak predictions as well as weak empirical support when specifying IMEC.

In addition, the latent variable model may be taken to imply the existence of a biological cause (e.g., Ackerman et al., 2005; Luciano et al., 2005; van der Maas et al., 2006) that constitutes the *g*-factor (E8). That this predicted phenomenon was never discovered reduces the explanatory coherence of the latent variable theory. Investigating how the discovery of such a biological basis for intelligence would change the explanatory relations of the two theories can tell us whether finding a biological correlate would constitute a critical experiment (e.g., Platt, 1964). In other words, an experiment that could discard the more coherent mutualism theory in favor of the latent variable theory. With IMEC we can test this by modeling a counterfactual world in which a common biological cause of *g* is discovered. To do so we assign a positive instead of a negative threshold to E8 and compare the theories again. Computing the explanatory coherence assuming a positive threshold on NE8 shows that finding a biological correlate to *g* would indeed constitute a critical test. After finding such a correlate the explanatory coherence of the mutualism theory would be .782, while the latent variable theory increases to .836, surpassing even the explanatory coherence of the mutualism theory in the configuration we started with. Thus, in this thought experiment, the latent variable explanation would be the preferable theory after discovering that it designates a biological feature of the human being.

The example illustrates how IMEC can compare psychological theories. We showed that the explanatory coherence of two theories can be compared with identify which of them constitutes a better explanation. In addition, by varying thresholds and weights we can account for higher or lower corroboration of phenomena. Finally, modeling counterfactual states of a theory (e.g., assuming a not yet discovered phenomenon to be discovered) can help us to identify critical experiments that would make a previously inferior explanation the better explanation. However, the example should also be considered a cautionary note as it shows how the conclusions derived from IMEC can depend on what are considered individual hypotheses. For example, one could make the case that when including the assumption of the mutualism model that growth can be described by a logistic model, it would be important to also include that the influence of intelligence on test scores is linear in the latent variable model. Determining inclusion and

Figure 8
Explanatory Relations for the Comparison Between the Positive Manifold Theory of Intelligence and Latent Variable Models of Intelligence



Note. Corresponding phenomena and propositions can be found in Table 1. The default edge weights are one; however, edge weights are split when multiple propositions are needed to explain a phenomenon. HM1 and HL1 have a strong negative connection of -4 . Thresholds for the phenomena are set to 2 and to -2 for the phenomenon E8. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

exclusion of explanatory hypotheses can indeed be considered a weakness of our model and we give more guidance about it in the discussion section.

Comparing Creationism and Darwin's Theory of Evolution

One intuitive approach to theory appraisal is selecting the theory that best explains the relevant phenomena.¹⁶ The complement of this intuition is that one should not abandon a theory until there is another theory that is better at explaining or explains more of the relevant phenomena. Thus, a case could be made that the coherence of an explanation should depend on the quality of the available alternative explanations.

One historical case, where a theory lost explanatory coherence because a better alternative explanation became available is when creationism was surpassed by the theory of evolution. As Thagard (1989) points out, Darwin's argument was explicitly comparative. In many places in the *Origin of Species* (Darwin, 1859) he points to facts that the theory of evolution explains while they were unexplainable by the previous idea that species have been created by God. It is also notable that creationism is a very simple theory—the only hypothesis it postulates is the existence of God. Therefore, creationism might seem attractive in light of parsimony. Only a model that considers different epistemic values will be able to refute creationism if a better alternative (evolution) becomes available. In the next section, we first reproduce an example from Thagard (1989) comparing creationism to the theory of evolution.¹⁷ Then we consider whether it was reasonable to believe creationism before the theory of evolution was proposed.

The explanatory relations can be found in Figure 9. Comparing the two theories with IMEC indicates that Darwin's theory is the superior explanation (.820) compared with creationism (.445).

However, in the case of the theory of evolution, one could argue that not all hypotheses should be considered but only the two core hypotheses of the theory. These are DH2 (Organic beings undergo natural selection) and DH3 (Species of organic beings have evolved). Therefore, we can also calculate the coherence of Darwin's theory as the mean explanatory coherence over only these two hypotheses, instead of all of Darwin's hypotheses. If we do so explanatory coherence of evolution increases even further to .999. This shows how explanatory coherence changes depending on the set of hypotheses considered elemental for the theory.

However, creationism provided an explanation for the origin of species long before Darwin conceived the theory of evolution. It is interesting to consider whether creationism was a coherent belief system at that time or whether it was always foolish to believe that species have been created by God.

To do so, we evaluate the creationist propositions in isolation. In other words, we do not model Darwin's theory and the resulting contradiction. Doing so shows that creationism would have had a much higher level of explanatory coherence before Darwin's theory. Before Darwin's theory emerged as a strong alternative the explanatory coherence of the creationist explanation was in fact .999. This high level of explanatory coherence may partly explain why people used to believe in creationism. It also illustrates that theories can lose coherence or be refuted once a better theory becomes available. Naturally, this example rests on the assumption that one considers the creationist explanation of the basic phenomena adequate; for those who did not accept this, the explanatory

¹⁶This approach is known in philosophy of science as *Inference to the Best Explanation* (IBE; e.g., Haig, 2009; Harman, 1965; Mackonis, 2013). As far as we know, none of the current attempts at formalizing IBE are without controversy, but there does not appear to be controversy about the intuition or the general principle.

¹⁷Unlike Thagard (1989) we do not model analogy.

coherence of creationism was low even in the absence of alternative theories. Thus, the example also shows that the dependence of the IMEC system on prior coding of whether or not propositions explain phenomena introduces a subjective element into theory evaluation. One can either accept this (rendering theory choice a partly subjective matter) or develop an objective account of explanation that can be used to determine whether or not a proposition truly explains a phenomenon.

Finally, we can also consider contradictions between theory and phenomena. For example, E5 (species become extinct) and E7 (forms of life change almost simultaneously around the world) can be considered directly in conflict with creationism.¹⁸ We can use IMEC to check whether the high individual explanatory coherence of creationism when evaluated in isolation is robust to these contradictions by adding contradictions between creationism and E5 as well as E7. This results in an explanatory coherence value of .39. This low coherence indicates that after discovering evidence that species become extinct and change, the theory of creationism was in peril (or at least its versions that could not account for this), which could have motivated the search for more explanatory coherent theories.

Discussion

In this article, we introduced a way to formally assess the quality of theories in psychology by implementing the theory of explanatory coherence using the Ising model. We show that our model smoothly integrates criteria of explanatory breadth, refutation, simplicity, and downplaying of evidence. In addition, by comparing two different models of intelligence, we show that IMEC is capable of comparing more complex theories and helps researchers to think through hypothetical scenarios and to identify critical experiments. Finally, the example comparing the theory of evolution to creationism illustrates that IMEC allows researchers to evaluate how the coherence of a theory changes when a better alternative is proposed. These different possible applications suggest that IMEC is a useful tool to facilitate theory evaluation.

In the above cases, thresholds and weights were set to the defaults because it is usually difficult to establish them objectively from the literature (Thagard, 1989). However, IMEC allows for changing these parameters in cases where researchers have empirical reasons to do so. Regarding the evidence for the phenomena, the thresholds should be scaled according to their strength of empirical support (e.g., Meehl, 1990). It is also important to consider how to scale the edge weights between different propositions and phenomena. The higher the edge weights between propositions and phenomena, the more strongly a theory is supported if there is support for the phenomenon. Based on different approaches for statistical inference, one can think differently about how to specify these edge weights. They could, for example, be scaled according to the probability of the phenomenon absent the theory because the posterior probability of a theory after a confirmed prediction depends upon this parameter (e.g., see Oberauer & Lewandowsky, 2019). Alternatively, information theory could ground the decision on the edge weights and the thresholds, for example, by trying to establish the distance between predictions and observations based on measures such as the Kullback-Leibler divergence (Kullback & Leibler, 1951).

The process of laying out the compared theories explicitly in terms of their hypotheses, explained phenomena, and the evidence for different phenomena might in itself already aid theory evaluation. First, weaknesses of a theory often become more apparent when formalizing it explicitly and the developed representation of the theory can be discussed with the scientific community. This should help to identify problems and possible modifications of theories already before evaluating them with IMEC. Second, scientists who have not developed a theory that specifies the explained phenomena concretely will notice this when attempting an IMEC implementation and can modify their theory to make more explicit predictions.

Other possibilities for theory evaluation are based on Bayes's theorem (Salmon, 2017) or Bayesian networks. While constructing a theory evaluation tool similar to IMEC based on Bayesian networks would certainly be an interesting possibility, it might in practice be difficult for researchers to specify the different prior probabilities and conditional probabilities required. For instance, it will be hard for a scientist to determine the conditional probability of the inheritance of mental qualities assuming animals do not have instincts. From this perspective, the simple structure of IMEC and its relatively transparent operation could be seen as strength.

Finally, IMEC's possible applications reach beyond the area of theory evaluation. Indeed a variety of explanatory processes can be understood in terms of explanatory coherence. Pennington and Hastie (1986) argue that legal decision making can be understood in terms of explanatory coherence and Thagard (1989) evaluates several cases of legal decision making using ECHO. Furthermore, ECHO was used to model psychological processes such as belief revision (Ranney & Thagard, 1988). Therefore, it should in principle be possible to extend IMEC to these areas.

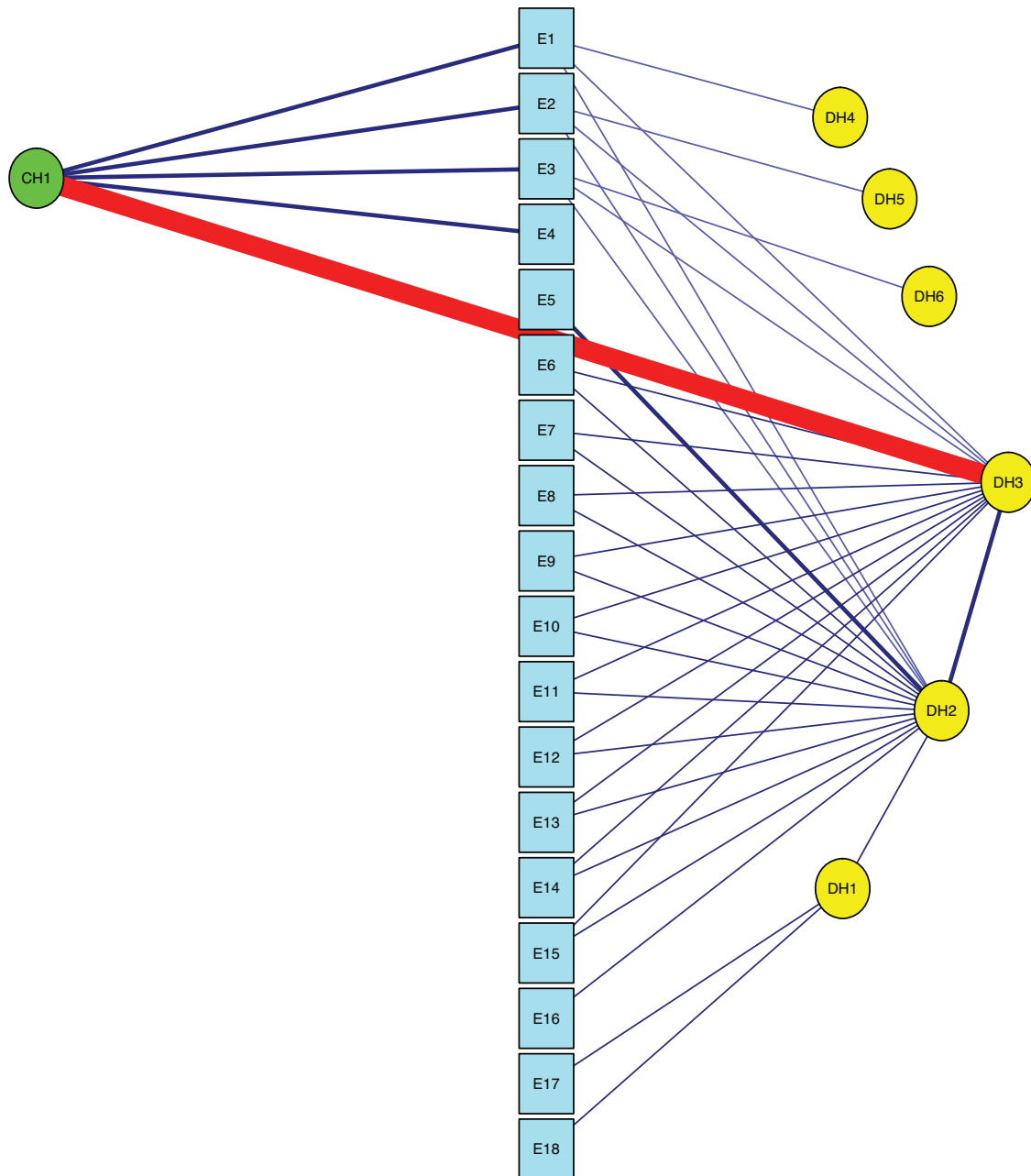
Limitations

As all models, our model has a variety of assumptions and limitations. First, the symmetry of explanatory connections. In practice, theories might be captured more precisely by directed networks as the Bayesian networks described above, even though these are more difficult to specify for researchers. Exploring a theory of explanatory coherence based on directed networks would be interesting future research. However, IMEC seems to trade off well between describing the complexity of theories and being simple enough for researchers to specify it in practice.

Second, while the principles of explanatory coherence seem suitable to compare a wide range of cases, they are only hypotheses themselves and, thus, open to revision. In other words, the principles that we based IMEC on may need to be changed again in the future. In addition, when assessing a theory researchers might sometimes want to incorporate criteria beyond explanatory coherence as incorporated in IMEC. For example—although explanatory breadth is already incorporated in IMEC—it seems conceivable that of two theories with the same explanatory coherence, researchers would prefer the one with more explanatory breadth. The reason for this could be that its wider set of predictions is

¹⁸ Although the creationist would likely argue this is not the case as God could create a large initial pool of species more of which go extinct over time, which is why these relations were not modelled in the main part.

Figure 9
Explanatory Network of the Comparison Between Creationism and Darwin’s Theory of Evolution



Note. The default weight is 1 but it is split if multiple propositions are needed to explain a phenomenon. The weight on the contradictions is -4 . Thresholds on the phenomena are all 2. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

more useful from an applied perspective. This shows how considerations beyond the principles of explanatory coherence as applied here can guide theory evaluation.

Third, a key weakness of using IMEC to compare different theories is that the parameter values need to be specified by the user. Although the default values seem to work well across a

wide range of cases; it is sometimes necessary to specify the edges and thresholds manually to incorporate weaker (stronger) evidence or weaker (stronger) predictions. In addition, determining the correct number of assumptions and phenomena is a difficult task. This process introduces researchers degrees of freedom and might allow a form of “theory-hacking,” where scientists

push the explanatory coherence of their own theory. Therefore, when using IMEC to settle theoretical debates in practice it is important to check the robustness of the conclusion by varying the parameter values. While much further research is needed, Thagard (1989) appealed to the following detailed maxim, which also applies to IMEC:

In analyzing the propositions and explanatory relations relevant to evaluating competing theories, go into as much detail as is needed to distinguish the explanatory claims of the theories from each other, and be careful to analyze all theories at the same level of detail. Following this maxim removes much of the apparent arbitrariness inherent in trying to adjudicate among theories.

Fourth, in some contexts theory and conceptualization scaffold each other, and it is difficult to construe them independently, or postulate that different theories manifest in identical phenomena (Danziger, 1985, 1993). For example, psychoanalysis and cognitive behavioral therapy might have different concepts of what counts as a valid measurement of depression (i.e., associations or projections vs. symptom checklist). However, while this might apply in specific cases, we believe that most phenomena, such as reaction times and error rates or the examples analyzed in this article, can likely be construed independently of the theories being evaluated. In addition, where this limitation applies theory comparison is also intractable with methods other than IMEC.

Fifth, IMEC is capable only of comparing specified theories. The process of generating explanatory hypotheses is not incorporated in IMEC at this point. In other words, finding a theory is still the responsibility of the human theorist. However, several computational algorithms for theory development have been proposed (Romdhane & Ayeub, 2011; Thagard & Nowak, 1988). These algorithms can infer hypotheses from data using certain inference rules. For example, an algorithm called PI (processes of induction) can infer explanatory hypotheses from simple statements (Thagard & Nowak, 1988). For instance, when instructed that sounds propagate and reflect PI will automatically create the concept of a *sound-wave*. Integrating such an algorithm with IMEC, for example, by evaluating how well the sound-wave theory predicts other observations about sound, could possibly facilitate smooth integration of theory development and theory appraisal.

Future Research

These limitations indicate directions for future research. First, it would be important to establish ways to derive the hypotheses and edge weights more systematically, for example, using text mining approaches to directly extract the explanatory relations from a paper. In addition, researchers could rely on the aforementioned computational algorithms for abduction (e.g., Romdhane & Ayeub, 2011; Thagard, 1988). Finally, severe testing, Bayesian inference and information theory might also guide the choice of edge weights as described before. This line of research could result in general knowledge about the nature of theories that would apply beyond IMEC.

Second, the original ECHO model has a variety of implications beyond the evaluation of scientific theories (Thagard, 1989, 1992). Therefore, it will be interesting to generalize IMEC beyond the realm of theory appraisal in psychology. Thagard (1989) applied

his original ECHO model to several cases of legal reasoning, which might also be a promising application for IMEC. In addition, IMEC could also be used to model human reasoning more generally. Ranney and Thagard (1988) show that ECHO could be used to model belief revision in naive physics. Coherence-based models have also been used to model cognition more broadly (Thagard, 2000; Van Rooij, 2008; Van Rooij et al., 2019). It should be possible to apply IMEC to model these kinds of problems with little adaptation.

Third, the Ising model shows interesting properties such as phase transitions (Barabási, 2016) depending on the value for the entropy parameter β . This parameter describes the heat of the network or the entropy of the model; in other words, if β is zero, the probability of each state is equal, whereas for high values of β , the network tends to settle in a single state. An interesting property of the Ising model is, that it shows a phase transition for high values of β but not for low ones. A phase transition describes a sudden change between the internal symmetry of the model (Solé, 2011); in the Ising model, this would be a change from an unordered (i.e., some magnets north pole up, others north pole down) to an ordered state (i.e., all nodes north/south up) of the system. Phase transitions have not been used in the present article and it would be interesting to consider the progression of research lines in terms of phase transitions. One could for example argue, that Popper's idea of falsification describes a phase transition (Popper, 1959), whereas Lakatos (1978) concept of progressive and degenerating research lines describes slower changes. Because the β parameter moderates the occurrence of phase transitions, the Ising model could theoretically incorporate both of these ideas, suggesting a possible unification of different theories in philosophy of science.

Conclusion

Theories are one of the most important tools of cumulative science (Lewin, 1943). Especially in psychology, there has been little progress in the area of formal theory development, despite the high attention that is paid to this topic (Borsboom et al., 2021). One reason for this might be that a systematic, computational approach to evaluating theories lacked in this field. By incorporating a theory of explanatory coherence in the Ising model we developed such an approach. We hope that this will help researchers to evaluate the quality of their theories and will aid in increasing the quality of theories in psychology.

References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. <https://doi.org/10.1037/0033-2909.131.1.30>
- Barabási, A. L. (2016). *Network science*. Cambridge University Press.
- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *The Philosophical Review*, *97*(3), 303–352. <https://doi.org/10.2307/2185445>
- Borsboom, D. (2013). *Theoretical amnesia*. <http://osc.centerforopenscience.org/2013/11/20/theoretical-amnesia/>
- Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766. <https://doi.org/10.1177/1745691620969647>
- Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion

- been overestimated? *Frontiers in Psychology*, 5, Article 823. <https://doi.org/10.3389/fpsyg.2014.00823>
- Dahlman, C., & Mackor, A. R. (2019). Coherence and probability in legal evidence. *Law, Probability and Risk*, 18(4), 275–294. <https://doi.org/10.1093/lpr/mgz016>
- Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review*, 123(1), 2–22. <https://doi.org/10.1037/a0039802>
- Danziger, K. (1985). The methodological imperative in psychology. *Philosophy of the Social Sciences*, 15(1), 1–13. <https://doi.org/10.1177/004839318501500101>
- Danziger, K. (1993). Psychological objects, practice, and history. In H. V. Rappard, P. J. Van Strien, L. P. Mos, & W. J. Baker (Eds.), *Annals of theoretical psychology* (pp. 15–47). Springer.
- Darwin, C. (1859). *On the origin of species*. John Murray.
- Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, 23(2), 105–132. [https://doi.org/10.1016/S0160-2896\(96\)90008-2](https://doi.org/10.1016/S0160-2896(96)90008-2)
- Duhem, P. (1954). *The aim and structure of physical theory*. Princeton University Press.
- Epskamp, S., Maris, G. K., Waldorp, L. J., & Borsboom, D. (2016). *Network psychometrics*. <https://arxiv.org/abs/1609.02818>
- Facon, B. (2004). Are correlations between cognitive abilities highest in low-IQ groups during childhood? *Intelligence*, 32(4), 391–401. <https://doi.org/10.1016/j.intell.2004.06.002>
- Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, 31(4), 271–288. <https://doi.org/10.1080/1047840X.2020.1853461>
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98(2), 254–267. <https://doi.org/10.1037/0033-295X.98.2.254>
- Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4), 789–802. <https://doi.org/10.1177/1745691620970585>
- Haig, B. D. (2005). An abductive theory of scientific method. *Psychological Methods*, 10(4), 371–388. <https://doi.org/10.1037/1082-989X.10.4.371>
- Haig, B. D. (2009). Inference to the best explanation: A neglected approach to theory appraisal in psychology. *The American Journal of Psychology*, 122(2), 219–234.
- Harman, G. H. (1965). The inference to the best explanation. *The Philosophical Review*, 74(1), 88–95. <https://doi.org/10.2307/2183532>
- Haslbeck, J., Ryan, O., Robinaugh, D., Waldorp, L., & Borsboom, D. (2021). Modeling psychopathology: From data models to formal theories. *Psychological Methods*. Advance online publication. <https://doi.org/10.1037/met0000303>
- Ising, E. (1925). Beitrag zur theorie des ferromagnetismus [Contribution to the Theory of Ferromagnetism]. *Zeitschrift Für Physik*, 31(1), 253–258. <https://doi.org/10.1007/BF02980577>
- Kellen, D. (2019). A model hierarchy for psychological science. *Computational Brain & Behavior*, 2(3–4), 160–165. <https://doi.org/10.1007/s42113-019-00037-y>
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <https://doi.org/10.1214/aoms/1177729694>
- Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge University Press.
- Lakens, D., Adolphi, F. G., Albers, C. J., Anvari, F., Apps, M. A. J., Argamon, S. E., Baguley, T., Becker, R. B., Benning, S. D., Bradford, D. E., Buchanan, E. M., Caldwell, A. R., Van Calster, B., Carlsson, R., Chen, S.-C., Chung, B., Colling, L. J., Collins, G. S., Crook, Z., . . . Zwaan, R. A. (2018). Justify your alpha. *Nature Human Behaviour*, 2(3), 168–171. <https://doi.org/10.1038/s41562-018-0311-x>
- Lewin, K. (1943). Psychology and the process of group living. *The Journal of Social Psychology*, 17(1), 113–131. <https://doi.org/10.1080/00224545.1943.9712269>
- Luciano, M., Posthuma, D., Wright, M. J., de Geus, E. J. C., Smith, G. A., Geffen, G. M., Boomsma, D. I., & Martin, N. G. (2005). Perceptual speed does not cause intelligence, and intelligence does not cause perceptual speed. *Biological Psychology*, 70(1), 1–8. <https://doi.org/10.1016/j.biopsycho.2004.11.011>
- Mackonis, A. (2013). Inference to the best explanation, coherence and other explanatory virtues. *Synthese*, 190(6), 975–995. <https://doi.org/10.1007/s11229-011-0054-y>
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2), 163–203. <https://doi.org/10.1037/0033-2909.109.2.163>
- Maier, M. (2021). IMEC (R package version 1.0). <https://cran.r-project.org/web/packages/IMEC/index.html>
- McPhetres, J., Albayrak-Aydemir, N., Mendes, A. B., Chow, E. C., Gonzalez-Marquez, P., Loukras, E., Maus, A., O'Mahony, A., Pomareda, C., Primbs, M. A., Sackman, S. L., Smithson, C. J. R., & Volodko, K. (2021). A decade of theory as reflected in psychological science (2009–2019). *PLoS ONE*, 16(3), Article e0247986. <https://doi.org/10.1371/journal.pone.0247986>
- McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *The American Statistician*, 73(Sup1), 235–245. <https://doi.org/10.1080/00031305.2018.1527253>
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46(4), 806–834. <https://doi.org/10.1037/0022-006X.46.4.806>
- Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1
- Meehl, P. E. (1992). Cliometric metatheory: The actuarial approach to empirical, history-based philosophy of science. *Psychological Reports*, 71, 339–339.
- Meehl, P. E. (2002). Cliometric metatheory: II. Criteria scientists use in theory appraisal and why it is rational to do so. *Psychological Reports*, 91(2), 339–404.
- Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 3(3), 221–229. <https://doi.org/10.1038/s41562-018-0522-1>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Nowak, G., & Thagard, P. (1992). *Copernicus, Ptolemy, and explanatory coherence*. <https://philpapers.org/rec/NOWCPA>
- Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26(5), 1596–1618. <https://doi.org/10.3758/s13423-019-01645-2>
- Peirce, C. S. (1958). *Collected papers: Science and philosophy, reviews, correspondence, and bibliography* (Vol. 7). Belknap Press of Harvard University Press.
- Pennington, N., & Hastie, R. (1986). Evidence evaluation in complex decision making. *Journal of Personality and Social Psychology*, 51(2), 242–258. <https://doi.org/10.1037/0022-3514.51.2.242>
- Platt, J. R. (1964). Strong inference. *Science*, 146(3642), 347–353. <https://doi.org/10.1126/science.146.3642.347>
- Popper, K. (1959). *The logic of scientific discovery*. Hutchinson & Co.
- Poston, T. (2021). *Explanatory coherence and the impossibility of confirmation by coherence*. <https://philpapers.org/archive/POSECA.pdf>
- Ranney, M., & Thagard, P. (1988). *Explanatory coherence and belief revision in naive physics*. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a201093.pdf>

- Robinaugh, D., Haslbeck, J., Ryan, O., Fried, E. I., & Waldorp, L. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, 16(4), 725–743. <https://doi.org/10.1177/1745691620974697>
- Robinaugh, D., Haslbeck, J., Waldorp, L., Kossakowski, J., Fried, E. I., Millner, A., McNally, R. J., van Nes, E. H., Scheffer, M., Kendler, K. S., & Borsboom, D. (2019). Advancing the network theory of mental disorders: A computational model of panic disorder. PsyArXiv. <https://doi.org/10.31234/osf.io/km37w>
- Romdhane, L. B., & Ayeub, B. (2011). An evolutionary algorithm for abductive reasoning. *Journal of Experimental & Theoretical Artificial Intelligence*, 23(4), 529–544. <https://doi.org/10.1080/0952813X.2010.505824>
- Salmon, W. C. (2017). *The foundations of scientific inference*. University of Pittsburgh Press.
- Sanbonmatsu, D. M., & Johnston, W. A. (2019). Redefining science: The impact of complexity on theory development in social and behavioral research. *Perspectives on Psychological Science*, 14(4), 672–690. <https://doi.org/10.1177/1745691619848688>
- Scheffer, M. (1997). *Ecology of shallow lakes* (Vol. 22). Springer Science & Business Media.
- Scheffer, M., Hosper, S., Meijer, M., Moss, B., & Jeppesen, E. (1993). Alternative equilibria in shallow lakes. *Trends in Ecology & Evolution*, 8(8), 275–279.
- Smaldino, P. (2019). Better methods can't make up for mediocre theory. *Nature*, 575(7781), 9. <https://doi.org/10.1038/d41586-019-03350-5>
- Solé, R. (2011). *Phase transitions*. Princeton University Press.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15(2), 201–293. <https://doi.org/10.2307/1412107>
- Thagard, P. (1988). *Computational philosophy of science*. MIT Press.
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12(3), 435–467. <https://doi.org/10.1017/S0140525X00057046>
- Thagard, P. (1991). The dinosaur debate: Explanatory coherence and the problem of competing hypotheses. In R. Cummins & L. Pollock (Eds.), *Philosophy and AI: Essays at the interface* (pp. 279–300). MIT Press.
- Thagard, P. (1992). *Conceptual revolutions*. Princeton University Press.
- Thagard, P. (2000). *Coherence in thought and action*. MIT press.
- Thagard, P. (2005). The emotional coherence of religion. *Journal of Cognition and Culture*, 5(1–2), 58–74. <https://doi.org/10.1163/1568537054068642>
- Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of Science*, 74(1), 28–47. <https://doi.org/10.1086/520941>
- Thagard, P., & Nowak, G. (1988). The explanatory coherence of continental drift. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 1988, pp. 118–126). Philosophy of Science Association. <https://doi.org/10.1086/psaprocbienmeetp.1988.1.192976>
- Tideman, E., & Gustafsson, J.-E. (2004). Age-related differentiation of cognitive abilities in ages 3–7. *Personality and Individual Differences*, 36(8), 1965–1974. <https://doi.org/10.1016/j.paid.2003.09.004>
- Vaidyanathan, U., Vrieze, S. I., & Iacono, W. G. (2015). The power of theory, research design, and transdisciplinary integration in moving psychopathology forward. *Psychological Inquiry*, 26(3), 209–230. <https://doi.org/10.1080/1047840X.2015.1015367>
- van der Maas, H. L., Dalege, J., & Waldorp, L. (2020). The polarization within and across individuals: The hierarchical ising opinion model. *Journal of Complex Networks*, 8(2), cnaa010. <https://doi.org/10.1093/comnet/cnaa010>
- van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- Van Rooij, I. (2008). The tractable cognition thesis. *Cognitive Science*, 32(6), 939–984. <https://doi.org/10.1080/03640210801897856>
- van Rooij, I., & Baggio, G. (2020). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, 16(4), 682–697.
- Van Rooij, I., Blokpoel, M., Kwisthout, J., & Wareham, T. (2019). *Cognition and intractability: A guide to classical and parameterized complexity analysis*. Cambridge University Press.
- van Rooij, I. (2019). *Psychological science needs theory development before preregistration*. <https://featuredcontent.psychonomic.org/psychological-science-needs-theory-development-before-preregistration/>
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, 13(4), 411–417. <https://doi.org/10.1177/1745691617751884>
- Woodward, J. (1989). Data and phenomena. *Synthese*, 79(3), 393–472. <https://doi.org/10.1007/BF00869282>
- Woodward, J. F. (2011). Data and phenomena: A restatement and defense. *Synthese*, 182(1), 165–179. <https://doi.org/10.1007/s11229-009-9618-5>

(Appendix follows)

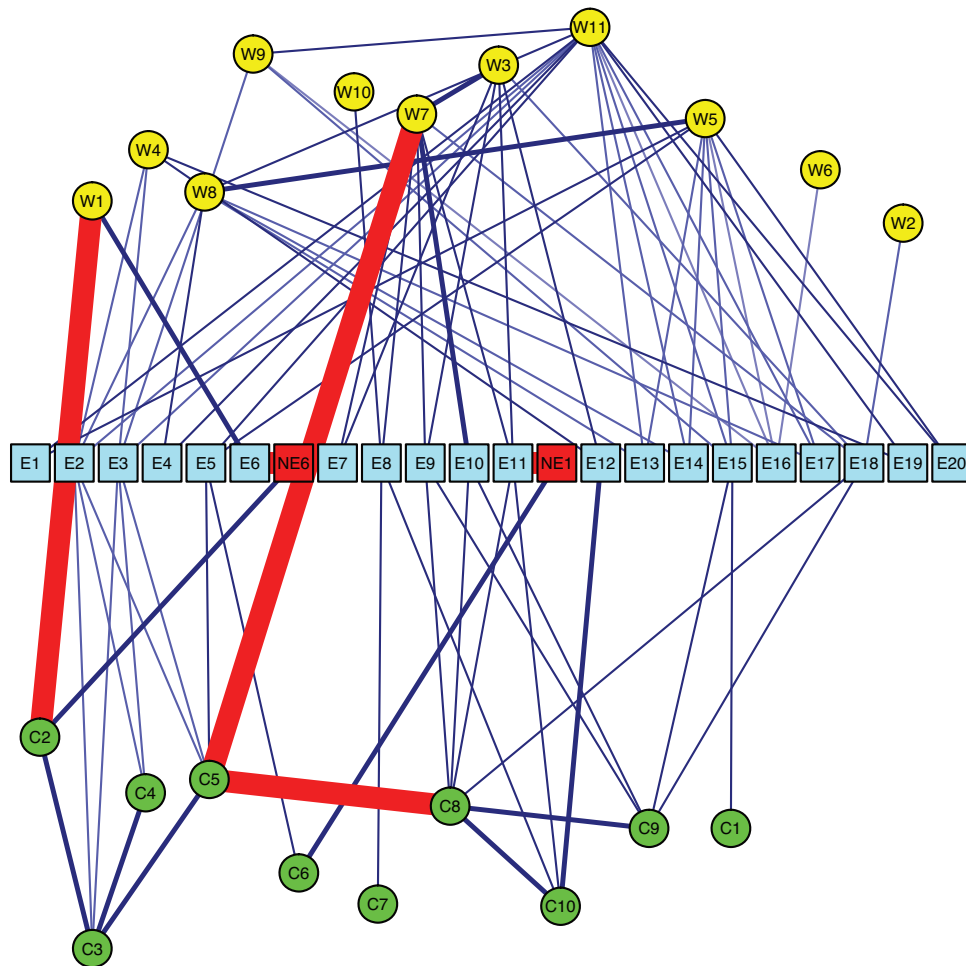
Appendix
More Applications of IMEC

Wegener’s Theory of Continental Drift

This example is the most complex theory comparison conducted with IMEC so far. Based on [Thagard and Nowak \(1988\)](#) we analyzed Wegener’s theory of continental drift. Wegener (as cited in [Thagard & Nowak, 1988](#)) was the first to propose a theory of continental drift based on the observation of strong similarity between the coastlines of Europe and North America. However, his theory was opposed by competing views, mainly the contractionist idea that shrinking of the

earth caused earthquakes. The explanatory network of this theory comparison is visualized in [Figure A1](#). Comparing the two theories with IMEC indicates that Wegener’s theory was indeed more explanatory coherence (.899) compared with contractionist theories (.534). However, [Thagard and Nowak \(1988\)](#) state that the inference depends on the input. Modeling the argument from papers written by Wegener’s opponents makes the contractionist theory more explanatorily coherent. Unfortunately, we cannot reproduce these analyses because they are not available in the article.

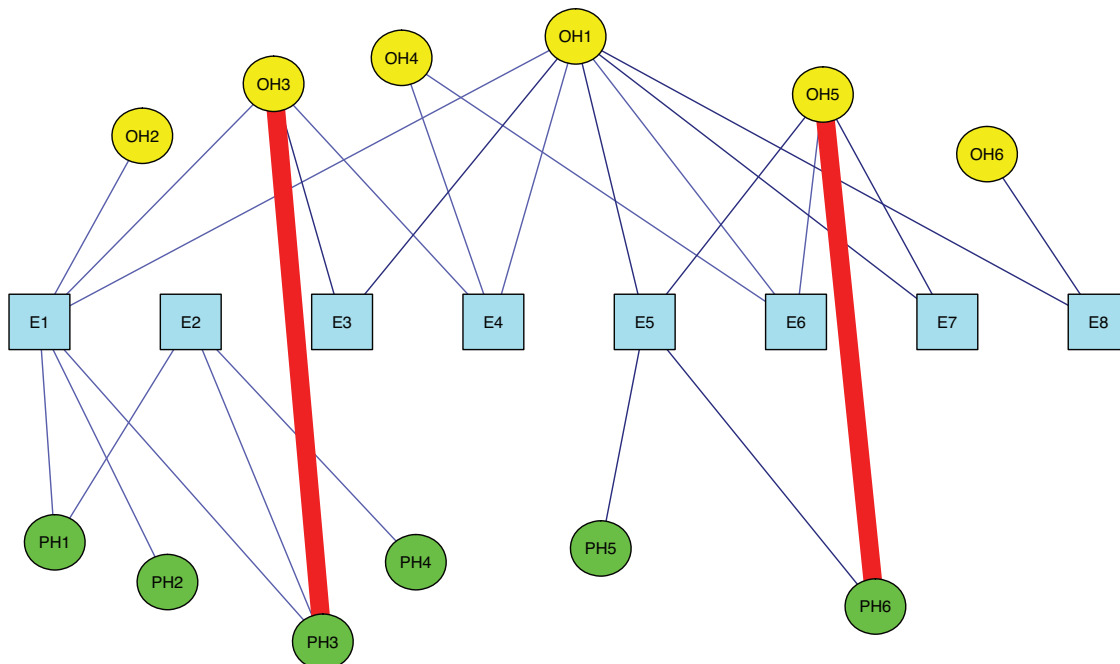
Figure A1
Explanatory Network of the Comparison Between Wegener’s Theory of Continental Drift and His Opponents



Note. The default weight is 1, but split if multiple propositions are needed to explain a phenomenon. The weight on the contradictions is -4 . Thresholds on the phenomena are all 2. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

(Appendix continues)

Figure A2
Explanatory Network of the Comparison Between the Oxygen and the Phlogiston Theory of Combustion



Note. The default weight is 1, but split if multiple propositions are needed to explain a phenomenon. The weight on the contradictions is -4 . Thresholds on the phenomena are all 2. H indicates the different explanatory hypotheses and E (for evidence) denotes the phenomena. See the online article for the color version of this figure.

Comparing the Oxygen and the Phlogiston Theory of Combustion

We now compare the phlogiston and the oxygen theory of combustion using IMEC. Thagard (1989) compares both theories and shows that the oxygen theory gives a better account for the process of combustion. However, because IMEC is also capable of evaluating individual theories, we can ask how explanatorily coherent the phlogiston theory was before Lavoisier proposed an alternative explanation based on oxygen. We now first reproduce the example from Thagard (1989) and then evaluate whether it was reasonable to believe the phlogiston theory before the explanation based on oxygen was proposed by evaluating this theory individually.

Up to the 18th century, the most important theory of combustion was the phlogiston theory proposed by Stahl (as cited in Thagard, 1989). This theory implies that combustion occurs due to burned bodies setting free phlogiston. In his essay “Reflexions sur le Phlogistique” Lavoisier (as cited in Thagard, 1989) proposed an alternative account of combustion based on oxygen. He claimed that combustion occurs because burned bodies are combining with oxygen in the air. Figure A2 shows the structure of explanations when comparing the two theories.

Based on this explanatory structure we can now compare the two theories using IMEC. As before the default weight is one, although it is split if multiple propositions are needed to

explain a phenomenon. In addition, we modeled strong contradiction with a weight of negative four on the contradictory propositions. Thresholds on the phenomena are all two. Implementing this model in IMEC shows that the oxygen theory’s explanatory coherence is .790 while the phlogiston theories’ explanatory coherence is only .543.

In this example, the phlogiston theory is downplayed because it contradicts the better-supported oxygen theory. However, it is interesting to also consider whether the phlogiston theory gave a reasonable account of combustion before Lavoisier developed the oxygen theory. To do this we can investigate the explanatory coherence of the phlogiston theory without modeling oxygen at all. Evaluating the theory in this way we find a higher explanatory coherence of .721. This shows that before Lavoisier proposed his theory of combustion, the phlogiston theory was a more appropriate explanation than after he proposed the theory. In other words, this illustrates how old explanations of a set of phenomena lose credibility if better explanations become available.

Received May 21, 2021

Revision received September 2, 2022

Accepted October 4, 2022 ■

Correction to “Comparing Theories With the Ising Model of Explanatory Coherence” by Maier et al. (2023)

In the article “Comparing Theories With the Ising Model of Explanatory Coherence,” by Maximilian Maier, Noah van Dongen, and Denny Borsboom (*Psychological Methods*, 2024, Vol. 29, No. 3, pp. 519–536, <https://doi.org/10.1037/met0000543>), the copyright attribution was incorrectly listed, and the Creative Commons CC BY 4.0 license disclaimer was incorrectly omitted from the author note. The correct copyright is “© 2023 The Author(s),” and the omitted disclaimer is below: Open Access funding provided by University College London: This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0; <https://creativecommons.org/licenses/by/4.0>). This license permits copying and redistributing the work in any medium or format, as well as adapting the material for any purpose, even commercially.

All versions of this article have been corrected.

<https://doi.org/10.1037/met0000705>