Informal Logic

How do Explanations Justify?
An Extended Scheme for Inference to the Best (Causal) Explanation

Petar Bodlović and Marcin Lewiński

Volume 44, Number 4, 2024

URI: https://id.erudit.org/iderudit/1116153ar
DOI: https://doi.org/10.22329/il.v44i4.8517

See table of contents

Publisher(s)

Informal Logic

ISSN

0824-2577 (print)
2293-734X (digital)

Explore this journal

Cite this article

Bodlović, P. & Lewiński, M. (2024). How do Explanations Justify? An Extended Scheme for Inference to the Best (Causal) Explanation. *Informal Logic*, *44*(4), 636–682. https://doi.org/10.22329/il.v44i4.8517

Article abstract

The paper presents an extended scheme for the inference to the best explanation (IBE). The scheme precisely treats the epistemic modifiers ("hypothetically," "plausibly," "presumably") of the inference, acknowledges its contrastive nature, clarifies the logical support between premises and conclusions (linked, convergent, and serial support), and introduces additional premises essential for inferring justified conclusions (especially those related to causal explanations and more demanding standards of proof). Overall, it advances the existing schemes for IBE in argumentation theory and treats IBE as a par excellence argumentative, rather than explanatory, form of reasoning.

How do Explanations Justify? An Extended Scheme for the Inference to the Best (Causal) Explanation

Petar Bodlović
Institute of Philosophy
Ulica grada Vukovara 54
10000 Zagreb
Croatia
pbodlovic@ifzg.hr

Marcin Lewiński
Nova Institute of Philosophy
Nova University of Lisbon
Campus de Campolide
1099-032 Lisbon
Portugal
m.lewinski@fcsh.unl.pt

Abstract: The paper presents an extended scheme for the inference to the best explanation (IBE). The scheme precisely treats the epistemic modifiers ("hypothetically," "plausibly," "presumably") of the inference, acknowledges its contrastive nature, clarifies the logical support between premises and conclusions (linked, convergent, and serial support), and introduces additional premises essential for inferring justified conclusions (especially those related to causal explanations and more demanding standards of proof). Overall, it advances the existing schemes for IBE in argumentation theory and treats IBE as a *par excellence* argumentative, rather than explanatory, form of reasoning.

Résumé: L'article présente en détail un schéma pour l'inférence vers la meilleure explication (IME). Le schéma traite précisément les modificateurs épistémiques (« hypothétiquement », « plausiblement », « vraisemblablement ») de l'inférence, reconnaît sa nature qui fait contraste, clarifie l'appui logique entre les prémisses et les conclusions (l'appui lié, convergent et sériel) et introduit des prémisses supplémentaires essentielles pour inférer des conclusions justifiées (en particulier celles liées aux explications causales et aux normes de preuve plus exigeantes). Dans l'ensemble, l'article fait progresser les schémas existants pour l'IME dans la théorie de l'argumentation et traite l'IME comme une forme de raisonnement argumentatif par excellence, plutôt qu'explicatif.

Keywords: abduction, argument scheme, causation, inference to the best explanation, justification, medicine, pragmatism, presumption.

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

1. Introduction

Explanations are ubiquitous. Due to common knowledge and usual practices, often we infer (most) plausible explanations spontaneously. For instance, after seeing a line in front of a restaurant on Saturday evening, we automatically conclude that these people want to have dinner or that it's a popular restaurant. Suppose, however, that COVID-19 infection rates rise dramatically over the next few weeks. What explains this? Virus mutations? Low vaccination rates? Are people not wearing masks? Not keeping distance? All of the above, but in different degrees? Sometimes, finding the best explanation becomes an inferential nightmare. What helps in such situations—besides having an expert, domain-specific knowledge—is the awareness of more abstract, semi-formal conditions that contribute to selecting the most plausible explanation. In this paper, we identify and elucidate logical relationships between such conditions.

Previous examples illustrate the famous kind of explanatory inference, namely, abductive reasoning. In the broadest sense, abduction amounts to "the whole process of generation, criticism, and acceptance of explanatory hypotheses" (Josephson and Josephson 1994, p. 8). For instance, in the medical context, a doctor may accept that the patient has hemolytic anemia because—given the patient's age, gender, medical history, etc.—having this disease explains her symptoms (e.g., the presence of panagglutinin on the laboratory test). Abduction is also common in science, law, and everyday life (Josephson and Josephson 1994; Douven 2021).

Surprisingly, a detailed theoretical study of inferring explanations from data started relatively recently, due to the works of C. S. Peirce and G. Harman.[1] Peirce, writing at the turn of the twentieth century, labeled this kind of reasoning an abduction, distinguished it from deduction and induction, and associated it with the initial stage of scientific inquiry where potential explanations are generated from data.

> Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value, and deduction merely

[1] Although, as most any problem in the theory of argumentation and reasoning, it can, arguably, be traced back to Aristotle; see Urbański and Klawiter (2018).

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

> evolves the necessary consequences of a pure hypothesis. (Peirce 1994, CP 5.171)

Harman (1965) focused on an evaluative (selective) rather than generative aspect of abduction (Wagemans 2016a; Yu and Zenker 2018) and labeled it the "inference to the best explanation" (henceforth, IBE). IBE instructs us to tentatively conclude that *H* is true if *H* provides the best explanation of data. Harman famously argued that all non-monotonic, ampliative reasoning is, in fact, IBE. That is because conclusions of such reasoning always rely—even if implicitly and thus potentially surreptitiously—on the idea that they are the best explanations one can infer from available evidence: "If we are adequately to describe the inferences on which *our knowledge* rests, we must think of them as instances of the inference to the best explanation" (Harman 1965, p. 94, emphasis added).[2] Ever since, disciplines such as the philosophy of science, epistemology, logic, artificial intelligence, and argumentation theory have studied explanatory considerations in connection to different modes of reasoning, scientific practice, justification, and causality.

Although philosophers offer extensive and rigorous analyses of IBE, their structural accounts of IBE often remain simplistic and vague. This results from their focus on abstract and, in a sense, fundamental theoretical issues, such as the importance of abduction for philosophical theories (e.g., pragmatism, scientific realism, hypothetico-deductivism) and scientific practices (see Peirce 1994, CP5, Book 1; van Fraassen 1980; Lipton 2004), or its logico-epistemic relations to other kinds of reasoning (see Peirce 1994, CP 5, Book 1; Harman 1965; Lipton 2004). Given such broader objectives, philosophers typically fail to enrich structural representations of IBE with their own theoretical insights. But there are other relevant objectives for which the study of IBE can and should be pursued. One such objective is a precise understanding of the IBE's normative structure (especially the one concerning causal reasoning) so that a detailed production and evaluation of various instances of IBE can be systematically undertaken. This is important for advancing the studies

[2] He illustrated this point by showing that IBE provides epistemic grounds for enumerative induction: we can justifiably conclude "All *A*'s are *B*'s" only if this generalization best explains data showing that all *observed A*'s are *B*'s.

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

of ordinary human reasoning, as well as computational argumentation. Our purpose here is to pursue such more pragmatic goals, but in the spirit of Harman's original analysis: our elaboration of the structure of IBE spells out the often-implicit elements of this ubiquitous form of reasoning, thus making IBE not only theoretically more precise, but also practically usable.

The paper proceeds as follows. Section 2 presents four accounts of the inferential structure of abduction (IBE) and identifies three problems: inadequate representations, structural vagueness, and unwarranted simplicity. In Section 3, we use concepts from philosophy, argumentation theory, and law to tackle these problems. We discuss different theoretical solutions available in the literature, choose the most promising ones, improve existing representations of IBE structures, and develop a new argument scheme for IBE (Section 4). In Section 5, we discuss some applications, implications, and qualifications of the new, extended IBE scheme. Section 6 summarizes our results.

2. Four structures and three challenges

IBE starts from data (e.g., a patient's laboratory tests show presence of panagglutinin), progresses via some explanatory connection (e.g., hemolytic anemia would explain the presence of panagglutinin), and ends with concluding that the proposition which, supposedly, best explains the data is true (e.g., "The patient has hemolytic anemia"). Let's start our analysis by presenting four schemes for abduction and IBE, proposed by influential authors: Peirce (1903, in Peirce 1994), Harman (1965), Josephson and Josephson (1994), and Lipton (2004).[3]

[3] Two important clarifications. First, authors use different propositional letters. For the sake of consistency and clarity, we substitute their letters with *H* (for the ultimate conclusion: Hypothesis, explanation, explanans, cause, diagnosis, etc.) and *E* (for the Explanandum, i.e., facts, data, or observations that require explanation). Second, strictly speaking, Harman and Lipton do not offer schemes, but schemes are directly reconstructable from their accounts of IBE. According to Harman (1965, p. 89): "In making this inference one infers, from the fact that a certain hypothesis would explain the evidence, to the truth of that hypothesis. In general, there will be several hypotheses that might explain the evidence, so one must be able to reject all such alternative hypotheses before one is warranted in making the

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

The surprising fact, *E*, is observed.
But if *H* were true, *E* would be a matter of course.
Hence, there is reason to suspect that *H* is true. (Peirce 1903, in 1994, CP 5.189*)*

Evidence *E*.
Hypothesis *H* would explain *E*.
H would provide a "better" explanation for *E* than would any other hypothesis.
Therefore, hypothesis *H* is true. (Harman 1965, p. 89)

E is a collection of data (facts, observations, givens).
H explains *E* (would, if true, explain *E*).
No other hypothesis can explain *E* as well as *H* does.
Therefore, *H* is probably true. (Josephson and Josephson 1994, p. 5)

Available evidence (*E*).
H, if true, would provide the best explanation of that evidence (*E*).
Therefore, *H* is (approximately) true. (Lipton 2004, p. 1)

These schemes are appropriate for introductory purposes, but they only express IBE's definition in structural terms, i.e., in terms of premises and a conclusion. In addition to spelling out the definition, a comprehensive argument scheme must be sensitive to broader theoretical considerations and do justice to several "surrounding" issues concerning IBE. Simplistic schemes are vague approximations and, as such, entail three distinct problems:

- *Inadequate representations*. Usually, some elementary units—i.e., premises and conclusions—lack accurate

inference. Thus one infers, from the premise that a given hypothesis would provide a 'better' explanation for the evidence than would any other hypothesis, to the conclusion that the given hypothesis is true." In Lipton's (2004, p. 1) words: "According to the model of Inference to the Best Explanation, our explanatory considerations guide our inferences. Beginning with the evidence available to us, we infer what would, if true, provide the best explanation of that evidence."

formulations. Propositions might come with inadequate epistemic or pragmatic modifiers (e.g., "probably" or "approximately") or lack modifiers. Note, for instance, that Harman does not qualify IBE's conclusion.

- *Structural vagueness*: Even if propositions are accurately formulated, logical connections between premises and conclusions are unclear. For instance, take two premises: "*H* explains evidence," and "*H* explains evidence better than any other hypothesis." Are these premises mutually independent units that jointly support the conclusion ("*H* is true"), or does the latter, somehow, logically presuppose the former? (Or even, as radical contrastivists would claim, the former *de facto* presupposes the latter: only best explanations in fact explain.) Simplistic schemes cannot answer such questions.
- *Unwarranted simplicity*: Even if connections between existing propositions are structurally clear, some premises, conclusions, and connections are simply missing. For instance, what makes *H* the best explanation of evidence? How are alternative hypotheses eliminated? How is selecting the best explanation related to one's personal interests or contextual goals? A complete scheme is wanting.

In the next section, we combine insights from philosophy, argumentation theory, and legal studies to develop a more accurate, detailed, and comprehensive representation of IBE's structure. First, philosophical work on contrastive explanations (Dretske 1972; Lipton 2004), and qualifiers like "hypothetically," "plausibly," and "presumably" (Godden 2017; Rescher 2006; Bodlović 2021) enables a more accurate representation of conclusion and premises. Second, in argumentation theory, the ongoing research on IBE's argument scheme (Walton 2001, 2006; Walton et al. 2008; Wagemans 2016a, 2016b; Yu and Zenker 2018; Olmos 2021), diagramming and argumentative patterns (Govier 2010) reduce structural vagueness. Finally, the philosophical research on causal explanations and interest-sensitivity (Lewis 1986; Lipton 2004) and legal scholarship on the standards of proof (Prakken and Sartor 2009; Macagno and Walton 2012) enable us to identify additional normative conditions and construct a stronger and more comprehensive version of IBE scheme.

For illustration purposes, we use examples from medical practice, IBE's natural area of application.

3. Developing an extended scheme for IBE

A diagnostic procedure begins with data. The doctor collects information about the patient's biography, medical history, and symptoms (based on the patient's testimony, physical examination, laboratory tests, etc.). Consider the following, real-life case from *casiMedicos*:[4]

> A 32-year-old woman with cerebral palsy from childbirth came to the emergency department after a few days of dark urine associated with an episode of high fever and dry cough. On admission, the CBC showed 16900 leukocytes/mm3 (85% S, 11% L, 4% M) ... In the biochemistry LDH 2408; bilirubin 6.8 mg/dl, … The morphological study of blood showed macrocytic anisocytosis with frequent spherocytic forms and polychromatophilia without blasts. The irregular antibody study is positive in the form of panagglutinin…

Suppose the doctor concludes that "The patient has hemolytic anemia associated with respiratory infection" because such a diagnosis provides the best explanation of the patient's symptoms. Two initial questions are essential for improving the existing IBE schemes: what is the nature of the doctor's conclusion, and how, exactly, should it be qualified? These interconnected questions are surprisingly difficult to answer.

3.1. What is IBE's conclusion H?

The question about the nature of the IBE's conclusion entails subquestions about its ontological and pragmatic features. Ontologically, it seems that, in many contexts, the conclusion must be factual. This seems obvious in the medical domain: a diagnostic procedure aims to discover and describe how the world is.[5] However,

[4] *casiMedicos* is a project where collaborators attempt to make various medical data more accessible to the public (see https://www.casimedicos.com/). For instance, in one part of the project, voluntary medical doctors explain examples from MIR (Medico Interno Residente) exams, necessary for practicing a medical specialization in Spain. We took our example from this data set.

[5] The factive nature of IBE's conclusion coheres well with Hempel's famous model of scientific explanation (see 1966, pp. 49-54). If (provisionally) translated

Josephson and Josephson (1994, p. 13) stress that the IBE's conclusion is, in an important sense, theoretical. IBE starts with observations but leads to the conclusion that *interprets* them. The doctor does not conclude "The patient has anemia" because she "sees" anemia, or because such diagnosis directly follows from observations in a trivial fashion, but because it provides a theoretical framework suitable for understanding data, i.e., (causally) explaining the patient's symptoms. In this sense, an acceptable conclusion of IBE is theoretical and factive (because it attempts to be true). However, in other contexts, IBE's conclusion can be acceptable without being factive. For instance, explanations which are strictly speaking false can nonetheless be acceptable if they are approximations that generate understanding, or idealizations that provide epistemic access to true beliefs.[6] So, IBE's conclusion is "*in principle*, theoretical or factual—*typically* mentioning either unobservable or merely unobserved entities, properties, and processes" (Olmos 2021, p. 136, emphasis added).

However, understanding the pragmatic nature of IBE's conclusion is itself a complex task, while being crucial for improving the existing IBE schemes. If the doctor concludes "The patient has anemia," what is the pragmatic, epistemic, or even discursive status of her conclusion? Is the diagnosis confirmed, or accepted?[7] And what kind of commitment does it represent? Cognitive commitments differ in strength and assume distinct normative requirements.

---

into Hempel's terms, the IBE's conclusion (e.g., "The patient has anemia") would represent a factual initial condition (an actual cause) that, together with some natural law, entails the explanandum ("The presence of panagglutinin on the patient's laboratory test, etc.").

[6] De Regt (2015, p. 3782) argues that, in some contexts, the existence of attractive forces might still explain gravitational phenomena, although Newton's theory of gravitation only approximates reality and Einstein's theory of general relativity refuted the existence of attractive forces. Elgin (2007, pp. 38-41) emphasizes that scientific explanations often include idealizations, although idealizations are false and do not purport to be true. Accordingly, fictive principles, such as an *ideal* gas law, might help us understand how some *actual* gas behaves. See Gaszczyk (2023) for further discussion.

[7] The notion of "confirmation" is typically used in probabilistic, Bayesian accounts of IBE. Its potential advantage is that it easily expresses "degrees of belief that fall short of full acceptance" (Lipton 2004, p. 63).

According to Peirce, abductive reasoning generates "explanations of phenomena held as hopeful suggestions" (1994, CP 5.196).

> Abduction merely suggests that something *may be*. Its only justification is that from its suggestion deduction can draw a prediction which can be tested by induction, and that, if we are ever to learn anything or to understand phenomena at all, it must be by abduction that this is to be brought about. No reason whatsoever can be given for it, as far as I can discover; and it needs no reason, since it merely offers suggestions. (Peirce 1994, CP 5.171)

Although this coheres well with Hempel's characterization of good scientific hypotheses as "happy guesses" (1966, p. 15), Peirce eventually gives more credit to abductive reasoning by portraying it as "intelligent guessing" (Peirce 1994, CP 6.530) that makes us "suspect" the conclusion is true (CP 5.189). But by being an intelligent guess, a hypothesis becomes "more than a possible explanation. It needs to be plausible as well" (Yu and Zenker 2018, p. 578). Thus, the modal qualifier "plausibly" leads us to the first candidate formulation of the IBE's conclusion, namely: "Plausibly, *H*" (e.g., "Plausibly, the patient has anemia").

This formulation works for generative abductive reasoning which produces the initial pool of hypotheses worth considering or testing, but is too weak for IBE that aims at selecting *the best* explanation. IBE's conclusion does not only represent a plausible but *the most* plausible explanation. It is the defeasible explanation that passed a demanding epistemic test by "winning" the selection process. Accordingly, IBE's conclusion is not (only) an intelligent guess and might even represent a piece of knowledge:

> If the abductive argument is strong, and if one is persuaded by the argument to accept the conclusion, and if, beyond that, the conclusion turns out to be correct, then one has attained justified, true, belief the classical philosophical conditions of knowledge… (Josephson and Josephson 1994, p. 16)

Although the second formulation of the IBE's conclusion—namely, "Knowingly, *H*"—pulls things in the right direction, it is too strong and restrictive. First, as Elgin (2007), de Regt (2015), and Gaszczyk

(2023) show, the best explanation might be non-factive in some epistemic contexts. Also, some explanations might be selected for non-epistemic reasons, such as harm reduction. Imagine the doctor must choose between the diagnoses *H* and *H** when *H* and *H** are, epistemically speaking, equally justified. Suppose, however, that *H* diagnoses the patient with a potentially deadly disease requiring immediate treatment, while *H** diagnoses her with a relatively harmless disease. Although the doctor cannot select the diagnosis on purely epistemic grounds, she can take precautionary measures, "err on the side of safety," and proceed as if *H* is true to protect human life.[8] In practical deliberation driven by non-epistemic goals, *H* might be the best explanation without being an epistemically justified belief.

This makes "presumably" the most suitable qualifier for the IBE's conclusion. There are several accounts of "presumption" (see Godden and Walton 2007; Lewiński 2017; Witek 2021). From a standard dialectical standpoint, presumptions shift the burden of proof, but from a logical viewpoint, presumptions are either default rules of inference (e.g., "If *H* would best explain why *E*, then, presumably, *H*"), or conclusions (e.g., "Presumably, *H*") inferred from such rules and basic facts (e.g., "*E* is true") (Rescher 2006, pp. 33-35). In this paper, we treat presumptions as conclusions of defeasible reasoning qualified with "presumably." "Presumably, *H*" indicates that due to epistemic, pragmatic, or, most usually, the combination of both epistemic and pragmatic considerations accepting *H* is the most reasonable way to proceed (until or unless evidence shows otherwise). So, the presumption is a "singulary" status (Freeman 2005, p. 26): while contrary conclusions may simultaneously appear plausible, only the most plausible one is a presumption. The modifier "presumably" entails "plausibly" or "defeasibly," but not *vice versa*.

Scholars usually distinguish cognitive (epistemic) from practical presumptions (Rescher 2006; Bodlović 2021). On the one hand, the concept of cognitive presumption fits perfectly with accepting *H* as the most plausible explanation. In Rescher's (2006) view, cognitive presumption "represents our most plausible candidate for truth" (p. 71). So, while cognitive presumption is the proposition that is,

[8] Namely, proceeding as if *H** is true when it is false (i.e., when *H* is true) might be fatal to the patient.

epistemically speaking, "more plausible than its potential rivals" (p. 39), IBE's conclusion usually includes an explanation that is, epistemically speaking, better than its potential rivals. Once we recognize that explanations can, indeed, generate presumptions, the correspondence becomes even more obvious. Namely, Rescher believes that a proposition can be presumed due to its "epistemic utility," i.e., because it would "if accepted, explain things that need explanation" (p. 47). In effect, when selected on epistemic ground, IBE's conclusion should be qualified with a cognitive (epistemic) version of the presumptive modifier: "Presumably$_{(c)}$, *H*."

On the other hand, the concept of practical presumption expresses accepting *H* as, pragmatically speaking, the most *suitable* explanation. Following the work of Ullmann-Margalit (1983) and Godden (2017), Bodlović (2021) suggests that *p* is a practical presumption "if an agent proceeds on *p* to promote a non-epistemic goal and, typically, to avoid greater harm in circumstances of uncertainty and pressure" (p. 289). This definition perfectly describes the doctor's reasoning in the previous case of practical deliberation: *H* is uncertain, the medical treatment urgent, and the doctor accepts an explanation *H* ("The patient has a deadly disease") to protect the patient's life. Accordingly, when selecting IBE's conclusion on pragmatic grounds, we should qualify it with a practical version of the presumptive modifier: "Presumably$_{(p)}$, *H*." Since our general scheme should be sensitive to both epistemic and pragmatic considerations, we might generalize the conclusion, as follows: "Presumably$_{(c/p)}$, *H*."

Someone may object that "Presumably$_{(p)}$, *H*" blurs the line between IBE and practical reasoning. If protecting a patient's life is the ultimate goal, concluding "The patient has a deadly disease" (*H*) is based on weighing practical costs rather than assessing *H*'s explanatory potential. In the context of epistemic uncertainty, the doctor might infer "The patient has a deadly disease" to avoid a costlier mistake, but still be convinced that "The patient has a harmless disease" (*H**) provides a better explanation of her symptoms. Accordingly, explanatory and practical considerations should remain separated.[9] Although this poses a genuine challenge for choosing

[9] We thank the anonymous reviewer for formulating this challenge and requesting further clarifications.

"Presumably$_{(p)}$" as an adequate modifier, linking IBE to practical rationality and cost management is common in the literature. For instance, Harman writes:

> There is, of course, a problem about how one is to judge that one hypothesis is sufficiently better than another hypothesis. Presumably such a judgment will be based on considerations such as which hypothesis is simpler, which is more plausible, which explains more, which is less *ad hoc*, and so forth. (Harman 1965, p. 69)

Arguably, choosing the simpler or less *ad hoc* hypothesis is a matter of practical or economic rationality, i.e., cost-benefit analysis. In Rescher's words:

> Simpler (more systematic) answers are more easily codified, taught, learned, used, investigated, and so on. […] It is the very quintessence of foolishness to expend greater resources than are necessary for the achievement of our governing objectives. […] The rational basis for preferring inductive simplicity lies in considerations of the economic dimension of practice and procedure rather than in any factual supposition about the world's nature. (Rescher 2006, pp. 126-127)

So, at least sometimes, explaining is governed by practical cost-benefit considerations that promote *epistemic* ends. However, scholars also recognize that *non-epistemic* cost-benefit considerations affect the explanation's quality. When discussing the nature of the abductive conclusion, Josephson and Josephson remark:

> Beyond the judgment of its likelihood, willingness to accept the conclusion should (and typically does) depend on:
> 1. pragmatic considerations, including the costs of being wrong and the benefits of being right
> 2. how strong the need is to come to a conclusion at all, especially considering the possibility of seeking further evidence before deciding. (Josephson and Josephson 1994, p. 14)

Walton (2001, 2006), Yu and Zenker (2018), and Olmos (2021) also acknowledge the relevance of pragmatic, cost-benefit considerations to IBE (mostly by referring to Josephson and Josephson). So,

assuming a purely epistemic account of explaining according to which practical considerations can never affect explanatory values appears too restrictive. Moreover, explanations sometimes promote epistemic goals of a very practical nature. For instance, explaining is more practically oriented when enhancing "understanding" (instead of "knowledge") since understanding is typically defined as a cognitive *ability* (see Grimm 2010; Hills 2016; Bodlović and Kudlek 2024). Finally, the "bestness" of explanation depends on several premises that have a pragmatic flavor (e.g., premises concerning the "standard of proof," or "pragmatic relevance" of a cause), as our extended IBE scheme will show. Hence, "presumably" is the most appropriate modifier because it can express the epistemic justification of IBE's conclusion, as well as the (potential) importance of pragmatic considerations.

Choosing the presumptive modifier has several implications. The most obvious implication is that some sort of qualification is needed, to begin with. Conclusion "Hypothesis *H* is true" (Harman 1965, p. 89) is simply too strong: "[F]or obvious reasons of fallibility, one should rather speak of the probable truth of the best hypothesis" (Dragulinescu 2016, p. 216). Josephson and Josephson also suggest that "*H* is probably true" (1994, p. 5). So, the natural question is: Why prefer "Presumably$_{(c/p)}$, *H*" over "Probably, *H*"?

Using "presumably" has several advantages. First, as explained above, some explanations are primarily selected for pragmatic reasons, and such reasons cannot be fully expressed in probabilistic, e.g., Bayesian terms. Second, even when focusing solely on epistemic considerations, probabilistic inference depends on having a big and representative sample and translating data into numerical values. Usually, however, numbers are unavailable, inadequate, and perhaps even unnecessary (Josephson and Josephson 1994, pp. 26-27; Dragulinescu 2016, pp. 211-213). Third, the highly probable explanation might not be the best. Lipton (2004, p. 59) offers an example: "It is extremely likely that smoking opium puts people to sleep because of its dormative powers … but this is the very model of an

unlovely explanation."[10] Some likely explanations such as this one are not informative and do not contribute to understanding the data.[11] Fourth, even if the likeliest explanation provides understanding, "probably" may entail "presumably." When *H* is a highly probable proposition, it usually becomes the most plausible proposition, i.e., a cognitive presumption: "The pivotal principle here authorizes the presumption of the probable" (Rescher 2006, p. 42).

To conclude, "probably" entails challenges that "presumably" does not, and when "probably" is adequate, typically, "presumably" is adequate, too.[12] So, in expressing the IBE's conclusion, "plausibly" is too weak, "knowingly" too restrictive, and "probably" too weak (as it allows uninformative explanations) and too restrictive (it does not allow uncertain, but pragmatically suitable explanations). In comparison, "presumably" seems adequate across all these scenarios.

Choosing "presumably" leaves us with one final lesson: it is incorrect to interpret the final conclusion of IBE as a hypothesis. To be sure, "hypothesis" is often used in everyday, loose sense. But to count as a hypothesis in a strict technical sense, *H* must have specific epistemic and pragmatic features. First, according to Peirce (1994,

---

[10] Conversely, one might select the best explanation that is highly unlikely. Sometimes, we must choose the best explanation among several unlikely, unpersuasive, bad explanations (see van Fraassen 1989).

[11] According to Lipton (2004), likeliness and loveliness reflect different goals of explanation, thereby providing distinct evaluation standards. Likeliness is oriented toward selecting a true, most warranted explanans in relation to the total available evidence. As a result, likely explanations are sensitive to defeaters, such as conflicting, competitive explanations. By contrast, loveliness is oriented toward selecting an explanans that increases our understanding of the data. However, understanding is not necessarily associated with total evidence and might be insensitive to epistemic defeaters. Like several other authors, Lipton remarks that Newtonian mechanics still offers a lovely explanation of the natural world despite being defeated and, consequently, unlikely. In Lipton's view, a plausible account of IBE must combine likeliness and loveliness. For instance, likeliness may assist us initially, while forming a set of plausible, potential hypotheses, while loveliness may assist us later on while choosing the best, presumed explanation from the set of initial hypotheses (see Lipton 2004, pp. 60-63).

[12] Rescher (2006, pp. 42-44) refers to the famous Lottery paradox to show that there are cases where low-probability propositions should, nevertheless, be presumed.

CP 5.196) *H* must be testable, i.e., entail empirical consequences enabling its confirmation or disconfirmation. Second, the goal of introducing *H* as a hypothesis is testing *H*: "The observable consequences derived from hypotheses are generated for the purpose of confirming or confuting the hypothesis itself" (Godden 2017, p. 501). Third, immediately after they are introduced, "hypotheses tend to have very little going for them, evidentially speaking" (Godden 2017, p. 501). They must fulfill some "minimal conditions … such as being consistent with all of the evidence," but are supposed to "accumulate [their] evidential merit along the way" (p. 501). This characterization makes it clear that the conclusion of successful IBE cannot be a hypothesis. At the final stage, *H* is already tested and, by surviving the test, becomes (epistemically) justified. Instead of testing whether *H* is true, we provisionally accept it and proceed as if *H* is true. So, *H* is a presumption, not a hypothesis.

3.2. What is IBE's explanandum E?

How should we formulate the basic premise triggering IBE or the so-called *explanandum E*?[13] In our medical example, *E* amounts to a set of symptoms ("The patient's laboratory tests show elevated LDH, presence of panagglutinin, etc.") the existence of which makes the conclusion ("The patient has hemolytic anemia associated with respiratory infection") acceptable. The conclusion (diagnosis), on the other hand, makes the explanandum (symptoms) understandable.

Scholars characterize explanandum in different terms. According to our introductory schemes, the conclusion of IBE is justified by some "evidence" (Harman 1965, p. 89; Lipton 2004, p. 1),

[13] We take this well-known term from Hempel (1966). The expression denotes the phenomenon that requires explanation or is being explained. An explanandum is usually uncontroversial, taken for granted, and does not require justification. In a dialogical context, it represents a common ground proposition, i.e., a claim, statement, event, fact, or a story that dialogue parties agree on, but may find puzzling and worthy of explanation (Walton 2005, 2011). For instance, both the doctor (the explainer) and the patient (the explainee) agree that the patient has dark urine, but are surprised by this anomaly and want to find an explanation of why this is the case. Whether an offered explanation must consist of true (Hills 2016), dialectically acceptable (van Eemeren and Grootendorst 2004, p. 72), or mutually understandable propositions (Walton 2005) is a difficult question, but, luckily, not essential to our present purposes.

"collection of data" (Josephson and Josephson 1994, p. 5), and "surprising fact" (Peirce 1994, CP 5.189). Furthermore, explanandum amounts to "observed findings" (Walton 2001, p. 144), "set of facts" (Walton 2001, p. 162; Wagemans 2016b, p. 102), "some event that has occurred" (Walton et al. 2008, p. 172), or some "shared or agreed upon ... data" (Olmos 2021, p. 137). Under scrutiny, such accounts of explanandum turn out to be both too narrow (restrictive) and too broad (permissive) for the sake of developing a general IBE scheme.

Previous characterizations apply to stereotypical cases of the explanandum. Typically, we need explanations when facing something unfamiliar, puzzling, or surprising, i.e., some anomaly that conflicts with our knowledge and expectations (Walton 2011). But IBE is not always motivated by a *surprising* fact. As Lipton remarks, "[t]he rattle in my car is painfully familiar, and consistent with everything else I believe, but while I am sure there is a good explanation for it, I don't have any idea what it is" (2004, p. 26). Furthermore, IBE might be triggered by an explanandum that, at time t_1, is not considered factual, or accepted as a shared belief. Imagine that, at t_1, "The patient has hemolytic anemia" is only a hypothesis, just one among many explanations that should be tested. In such circumstances, it is entirely natural to ask "What would explain hemolytic anemia?" and consider potential causes of this disease.[14] In other words, it is entirely natural to seek an explanation for a hypothetical statement. Consequently, previous characterizations are too narrow: sometimes, the explanandum is not surprising, factual, or agreed upon.

More importantly, the previous formulations hide the explanandum's contrastive nature. According to Lipton (2004, p. 33), the phenomenon that needs explaining "consists of a fact and a foil." In our example, the fact is the set of symptoms and the foil is some contrast class, explanatorily relevant in a given context.[15] For instance,

[14] Note that discussing causal explanations of hypothesis *H* is directly relevant for testing *H*. Suppose that having some viral infection (*H'*) often causes hemolytic anemia (*H*). Hypothesis *H'* may motivate the doctor to perform additional tests the results of which will, then, either increase or decrease the plausibility of "The patient has hemolytic anemia."

[15] What makes a contrast class "explanatorily relevant"? Or, in the words of one of the reviewers who pushes us on this point, "where does it come from?" We justify more precisely the relevance of contrast classes in pragmatic terms in Sec.

explaining why "The patient has a high fever, instead of no fever" differs from explaining why "The patient has a high fever, instead of low-grade fever;" and explaining why "The patient has a dry cough, instead of no cough" differs from explaining why "The patient has a dry cough, instead of wet cough," etc. Indeed, any proposition can have a different contrastive focus: "I *sold* my typewriter to Clyde" differs from "I sold my typewriter to *Clyde*" (Dretske 1972, p. 411). Recognizing contrastive differences is essential because they set normative boundaries. For instance, if the explanandum comes down to "*selling* instead of *giving* my typewriter," then "I needed the money" provides a good explanation. By contrast, if the explanandum comes down to "selling my typewriter to *Clyde*, instead of *Alex*," then something like "he called first" does the explanatory job, while the money explanation does not. Contrast classes represent the first, initial criterion for ruling out inappropriate explanations.[16]

To conclude, the current formulations of the explanandum are imprecise and broad. To narrow them down, we must make explicit that explanandum has a particular focus or contrast class: it does not amount to, e.g., "Surprising fact *C*," but "Surprising fact *C*, instead of *C**"; not "Data *D*," but "Data *D*, instead of *D**", etc. Also, as argued above, we must formulate an explanandum so that it can

3.5, while discussing the contrastivity of causal hypotheses / explanantia. The arguments given there hold for explananda too, but let us briefly touch on the explanandum-specific concerns. First, explananda are socially constructed via explanatory dialogue, not just "observed" or "found" (see Rohfling et al. 2021): they depend on the interests, motivations, and knowledge of both the explainee and the explainer. Second, when a contrast class is not stated explicitly, it can be reconstructed by taking into account mutual knowledge, common sense, prior talk, explainee's social role, speakers' interests, etc.—all of which generate some default expectations, which provide our "foil." Take a simple fact that "John is home". Why is he? It becomes a properly formulated explanandum to be (causally) explained *only after* we consider some contrast class. "Because he's sick" works as a cause for the contrast class "John is at work," "because he's much better now" for "John is in the hospital" and "because his flight was cancelled" for "John is travelling." Without such "foils" our attempted explanations are unguided and take us everywhere or, most likely, nowhere.

[16] For instance, an explanation that works for many contrast classes is usually trivial, i.e., too general for relevant contextual purposes. In some sense, "The patient was born" explains her symptoms, conjoined with all relevant contrast classes, but this is hardly an explanation the doctor is looking for.

include unsurprising facts and non-factual statements. So, instead of facts, data, observations, etc., perhaps we should be talking about some contrastive proposition (or set of propositions) that requires explaining: "*E*, instead of *E**".

So far, we addressed the problem of inadequate representations by formulating an IBE's conclusion and explanandum more precisely. Next, we also address the problems of structural vagueness and unwarranted simplicity.

3.3. How does IBE move from E to H?

At this point, our IBE scheme looks like this:

> *E*, instead of *E**. (e.g., high fever, instead of low-grade fever, etc.)
> Therefore, "Presumably (c/p), *H*. (e.g., Presumably, hemolytic anemia associated with…)

But what glues the two together? The scheme is enthymemic, and an implicit major premise (or: connection premise, presumptive rule, warrant) enabling the inference from "*E*, instead of *E**" to "Presumably$_{(c/p)}$, *H*" should be made transparent. The simplistic introductory schemes seem to suggest that the connection premise is explanatory. So, let's provisionally supplement our scheme with explanatory considerations found in Harman's (1965), Josephson and Josephson's (1994), and Lipton's (2004) schemes. To reduce structural vagueness associated with the standard form of argument presentation, we use argument diagrams that make distinct kinds of patterns and supports—e.g., linear (or serial), linked, convergent, divergent argument/support (Govier 2010)—more apparent.[17]

[17] Distinguishing between these inferential patterns entails a number of theoretical challenges (see Yu and Zenker 2022), but using diagrams and appealing to standard types of inferential support is still superior to presenting complex argument schemes in the standard form. That is, diagrams and inferential patterns allow us to better *approximate* logical relationships and provide us with preliminary tools to discover, and vocabulary to discuss "hard cases," such as structurally ambiguous inferences.

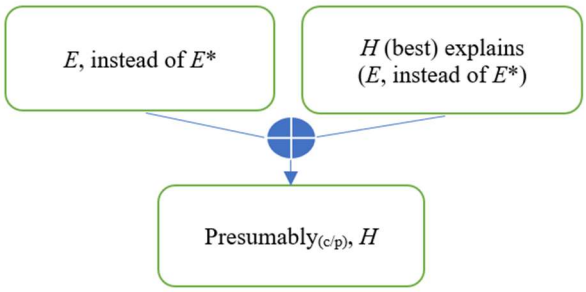

Diagram 1. *Simplistic schemes as linked argument*

This diagram depicts the linked argument where premises justify the conclusion "only when they are taken together; no single premise will give any support to the conclusion without the others" (Govier 2010, p. 37). The diagram seems plausible: explanatory premise "works together" with explanandum to justify "Presumably $_{(c/p)}$, *H*."

However, simplistic schemes jump to explanatory considerations too quickly. Let us take a step back, focus on the relation between *E* and *H* and, before proceeding any further, supplement our scheme with an implicit premise we are undeniably committed to: "If *E* (instead of *E**), then, presumably$_{(c/p)}$, *H*." We cannot justifiably infer *H* from *E* without accepting this conditional.[18]

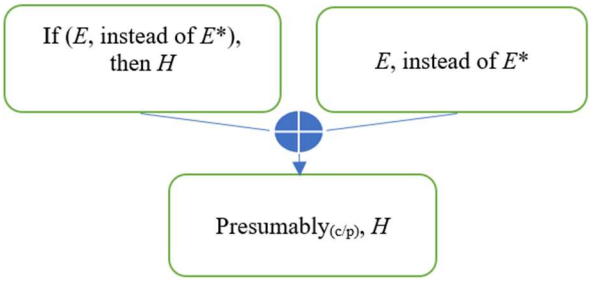

Diagram 2. *Logical minimum*

By adding the conditional, we reconstruct the argument in line with the "logical minimum" (van Eemeren and Grootendorst 2004, p. 117). Such a strategy appears trivial but helps to locate explanatory

[18] Suppose someone says: "I find inferring *H* from *E* acceptable, although the proposition 'If *E*, then *H*' is unacceptable (or false)." Without further clarification, this seems contradictory.

premises within IBE's overall structure. Since the conditional suffices to bridge the logical gap between *E* and *H*, treating explanatory considerations as an additional premise of a linked argument (as in Diagram 1) makes them logically redundant. Since explanations are the driving force behind IBE, this consequence is unacceptable.

The solution is clear: explanatory considerations justify the conditional. From the (descriptive) psychological perspective, believing that "*H* explains *E*" seems to motivate our *reasoning* from *E* to *H*. From the epistemic viewpoint, *H* being the (best) explanation of *E* justifies our *inference* from *E* to *H*. The latter interpretation coheres with Harman's (1965) well-known thesis that IBE, in fact, justifies enumerative induction: the inference from "All observed *A*'s are *B*'s" (*E*) to "All *A*'s are *B*'s" (*H*) is justified only if *H* best explains *E*.[19] Argumentation scholars mostly recognize that "our explanatory considerations guide our *inferences*" (Lipton 2004, p. 1, emphasis added). For instance, in Wagemans' IBE scheme, "H_i is the best explanation of *E*" supports "If it is observed that *E* is the case, we may assume that H_i is true" (2016a, p. 46). Similarly, in Olmos' IBE scheme "'*H* explains *D*' … is taken as a reason to support that '*D* justify *H*'" (2021, p. 142).[20] So, the explanatory premise justifies the conditional (rather than the conclusion *H*) in a "linear sequential pattern" (Govier 2010, p. 37).

[19] If the explanandum "All observed ravens are black" is better explained by some other phenomenon, e.g., biased sample (e.g., excluding ravens that are not black), or compromised perception (e.g., a color-blind agent), then enumerative induction is unreliable.

[20] However, neither Olmos' nor Wagemans' scheme is sensitive enough to contrast classes and modifiers.

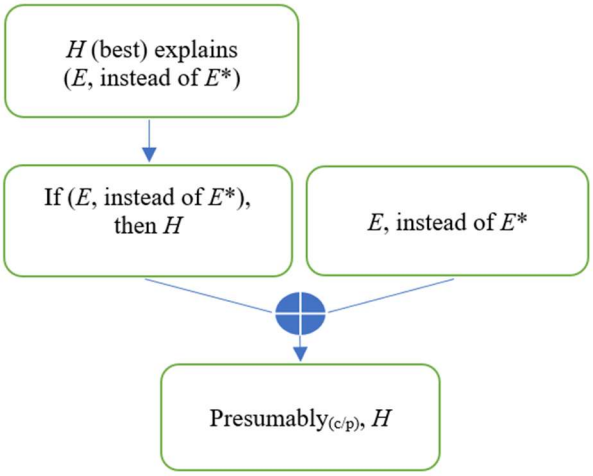

Diagram 3: *IBE: ultimate justificatory sequence*

However, there is a dilemma concerning the conditional's formulation. Among the four initial, simplistic schemes, only Peirce's includes the conditional, but formulates it differently: "If *H* were true, *E* would be a matter of course" (1994, CP 5.189). When constructing the scheme for the "argument from effect to cause,"[21] Walton et al. (2008) offer a similar formulation, whereby the cause (*H*) is the antecedent and the effect (*E*) the consequent of a conditional, but immediately reject it for two reasons. Firstly, in the case of IBE, such "cause to effect" formulation—e.g., "Generally, if *H* [cause] occurs, then *E* [effect] will (might) occur" (p. 170)—leads to the fallacy of affirming the consequent.[22] Secondly, and more importantly, the "cause to effect" formulation is misleading since it naturally reflects prediction. Namely, "If *H*, then *E*" seems to reflect reasoning that proceeds from cause (explanation) to effect (explanandum), while IBE is essentially "based on a retroduction" proceeding "from the observed data to a hypothesis about the presumed cause of the data"

[21] There is an obvious connection between this kind of argument and IBE since inferring the cause (e.g., diagnosis) comes down to inferring the proposition that (best) explains the effect (e.g., symptoms).
[22] In our opinion, this rejection of the "cause to effect" formulation might rest on a category mistake since affirming the consequent is a fallacy of deductive inference, while IBE is non-deductive.

(p. 171). Accordingly, we should prefer retroductive ("effect to cause") rather than predictive ("cause to effect") formulation.

3.4. The quality of explanation: the standard of proof

But on what grounds, exactly, is such retroductive conditional justified? Initial, simplistic schemes mention two distinct (but interconnected) explanatory premises that, so far, we have been treating as one.

The first, *comparative* explanatory premise, lies at the heart of IBE. Lipton (2004, p. 1) offers the most straightforward formulation: "[*H*] would, if true, provide the best explanation of that evidence [*E*]." Harman (1965) and Josephson and Josephson (1994) offer equivalent formulations. Going back to our medical example, this means that the clinician is justified to infer "The patient has hemolytic anemia associated with respiratory infection" (*H*) from "The patient has high fever (instead of…), dry cough (instead of…), elevated LDH (instead of…), etc." (*E*, instead of *E**) because diagnosis *H* would provide the best explanation of the patient's symptoms *E*. But how is *H*, then, qualified at this stage? Since *H* represents the most plausible explanation (*explanans*), according to our terminology from Section 3.1, *H* should be *presumed as an explanation*. Therefore:

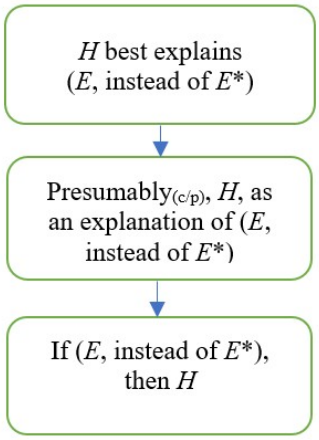

Diagram 4: *IBE: weak (basic) formulation*

This is an improvement in comparison to the schemes insensitive to modifiers that do not track how *H*'s epistemic status changes at different points or stages of justification. However, this is still IBE's *weak (basic) formulation*. Namely, even if hemolytic anemia is a presumed explanation, the inference from the patient's symptoms to "The patient has hemolytic anemia" may be poorly supported. As van Fraassen's (1989) "bad lot objection" suggests, *H* could be selected from the set of really bad explanations and, consequently, be comparatively the best without being epistemically satisfactory.

This brings us to the second, *non-comparative* explanatory premise. Namely, in addition to being the best available explanation, *H* must correctly explain "*E*, instead of *E**". Harman's and Josephson and Josephson's schemes recognize this condition. The latter, for instance, states: "[*H*] explains [*E*]" (1994, p. 5). Although the underlying message is clear, the latter formulation is unfortunate since "*H* explains *E*", if taken literally, seems presupposed by a comparative premise. "*H* best explains *E*" presupposes that "*H* explains *E*," so the two conditions are not logically independent as they should be. Walton improves Josephson and Josephson's scheme precisely in this respect by stating: "[*H*] is a *satisfactory* explanation of [*E*]" (2001, p. 162, emphasis added). Translated in our terms, "satisfactory explanation" becomes a "sufficiently plausible explanation." As far as the pragmatic (epistemic) status is concerned, even as a sufficiently plausible explanation (in a non-comparative sense) *H* is merely a hypothesis (albeit a reasonable one), since it still must be tested against other plausible explanations. So, in the IBE's standard formulation, comparative and non-comparative premises work together and provide linked support to explanation *H* being a *moderately strong presumption*.[23]

[23] For the general account of what contextual, justificatory, and deontic factors determine the strength of presumption, see Bodlović (2022).

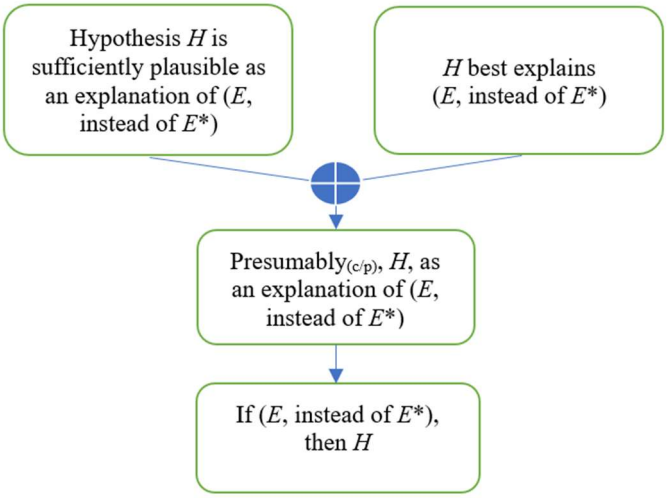

Diagram 5: *IBE: moderate (standard) formulation*

But even this might not suffice. Suppose two explanations *H* and *H** are both sufficiently plausible, and *H* is only slightly better than *H**. For instance, imagine that hemolytic anemia and marrow aplasia both explain the majority of a patient's symptoms, and hemolytic anemia explains only one (inconclusive) symptom more.[24] In this case, *H* is non-comparatively plausible, comparatively the best, but still relatively weak explanation: our acceptance of *H* instead of *H** rests on *H* being able to explain one (inessential) symptom. So, to be an exceptionally strong presumption, *H* must be at the particular "epistemic distance" from the second-best hypothesis *H**. Such normative condition is not acknowledged in current IBE schemes but is standardly recognized in the legal scholarship on the standard of proof.

[24] Another advantage of using modifiers "plausibly" and "presumably" is that they can easily express the kinds of cases where contradictory statements have high epistemic values: both *H* and *H** (i.e., not-*H*) can be highly plausible hypotheses. A Pascalian conception of probability (which is the dominant model today) has difficulties in expressing this because the sum of probabilities of contradictory statements must amount to 1. Put simply, if *H*'s probability value is high (say, 0.89), then non-*H*'s probability value must be low (say, 0.11) (see Woods, Irvine and Walton 2003, pp. 298-300).

Legal scholars distinguish between four proof standards. In terms of IBE, first, the explainer satisfies the weakest *scintilla of evidence* standard (SE) by (a) offering some undefeated explanation *H* (however weak this explanation may be). Second, she fulfills the moderate *preponderance of evidence* standard (PE) if she satisfies SE, and (b) *H* is more plausible than the strongest alternative explanation *H**. Third, the explainer satisfies the strong *clear and convincing evidence* standard (CCE) if she satisfies PE, but in addition (c) the plausibility of *H* exceeds some threshold α, and (d) the difference between the plausibilities of *H* and *H** exceeds some threshold β. Finally, she satisfies the highest *beyond a reasonable doubt* standard (BRD) if she fulfills CCE, and, simply put, (e) the epistemic distance between *H* and *H** is particularly large, i.e., threshold β is exceptionally high (see Prakken and Sartor 2009; Macagno and Walton 2012, p. 275; Bodlović 2022, p. 102). So, *H* would become a plausible explanation by satisfying SE; (weakly) presumed explanation by satisfying PE; strongly presumed by satisfying CCE; and, finally, an exceptionally strong presumption by satisfying BRD.[25]

Higher standards of proof (CCE and BRD) introduce the premise of *explanatory distance* that—linked with premises expressing plausibilities of *H* and *H**—supports the new, stronger formulation of a comparative premise: "*H* is a significantly better explanation of '*E*, instead of *E**' than *H**". Then, in the next inference step, the comparative premise—linked with the assumption that *H* and *H** represent the strongest explanations available—justifies strongly

[25] The choice of the appropriate standard of proof depends on the contexts, interests, and practical stakes. For instance, in a criminal trial, the prosecutor must prove guilt beyond a reasonable doubt because the freedom (or even the life) of the accused is at stake. Differently put, the hypothesis "John is guilty of murder" must explain the available evidence (bloody gloves, fingerprints, etc.) beyond a reasonable doubt in order to count as the best explanans. By contrast, in civil cases, attorneys must satisfy only the moderate preponderance of evidence standard because the cost of committing a mistake is lower (e.g., the defendant usually must pay some money, but won't spend a life in jail) (Prakken and Sartor 2009). Accordingly, IBE cannot easily be separated from the non-epistemic values and cost management: pragmatic factors determine the standard of proof, and the standard of proof determines "bestness." For more reasons showing that pragmatic cost-management also underlies scientific (i.e., genuinely epistemic) explanations, see Wilholt (2009).

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

presuming *H* as an explanation of "*E*, instead of *E**". Diagram 6 shows how these advanced comparative considerations fit in the new, strong formulation of IBE's structure.

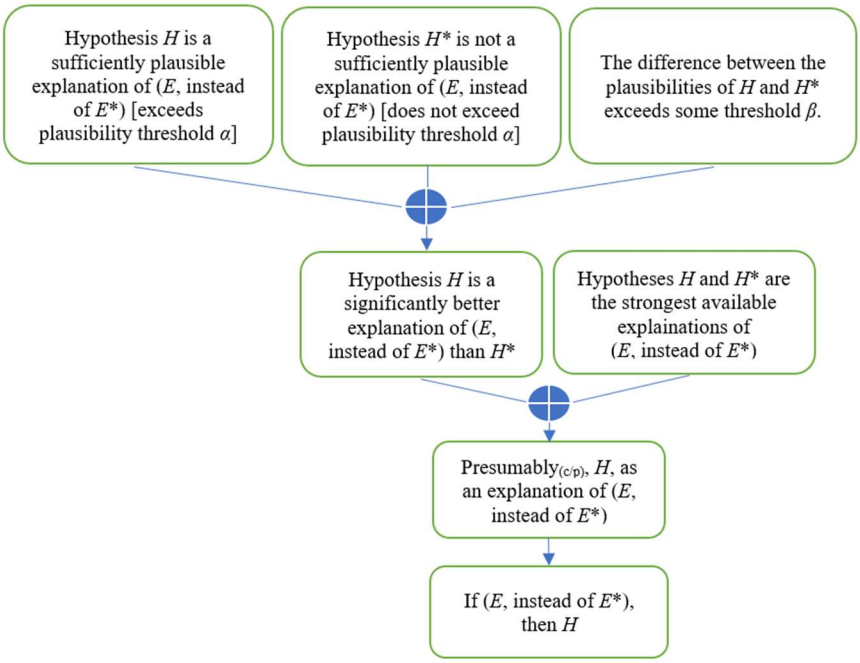

Diagram 6: *IBE: strong (new) formulation*

3.5. Accepting the sufficiently plausible (causal) hypothesis

Next, what makes an explanation *H*, unlike *H**, sufficiently plausible? This is an extremely difficult question since many factors determine plausibility and might pull in opposite directions.[26] In addition, explanations come with their own set of plausibility conditions. Namely, *H* must promote some unique explanatory virtues to be plausible as an explanation, i.e., to provide "understanding" (Lipton 2004; Walton 2005, 2011; Grimm 2010) or "knowledge why" (Hills 2016). Wagemans (2016b, p. 106) offers an extensive list of

[26] For instance, reliability of the source (perception, memory, expert testimony, etc.), evidential support, the absence of defeaters, coherence with existing knowledge, probability, normality, uniformity, simplicity, etc. (see Rescher 1976, pp. 6–7; 2006, pp. 38-44; Freeman 2005, p. 41).

explanatory virtues (usually discussed in the philosophy of science): accuracy of explanation, scope, genericity, fruitfulness, explanatory force, subsumptive power, refutability, empirical content (testability, observability), coherence, consistency, simplicity, elegance, parsimony, consilience (convergence of evidence), etc. However, some explanatory virtue might be more relevant to one kind of explanation (e.g., causal) than the other (e.g., statistical or logical explanation).

In this paper, we focus on the *scope* of *causal explanations*. This is a difficult choice because causality is hard to define. For instance, already in ancient times, Aristotle distinguishes between "four causes" (material, formal, efficient, and final) and corresponding types of explanation. Our present analysis assumes the most commonsensical, usual understanding of causation that, roughly, corresponds to Aristotle's efficient cause.[27]

> An event is considered the cause of another event if it precedes the caused event and is connected with it in a regular and consistent fashion. As such then the event is explained in terms of its antecedent cause. For example, if an organism is exposed consistently to a bacterium before developing a specific disease then the bacterium is said to cause the disease and serves as the principal etiological agent for explaining the disease. (Marcum 2008, p. 140)

So, to improve the IBE scheme, we assume that the explanations' purpose is acquiring knowledge about causal structures (Lewis 1986) and that *H*'s (diagnosis') plausibility depends on how many features of explanandum *E* (symptoms) *H* explains (without, also, explaining features that are not detected).[28] So, on what grounds can we accept *H* as a sufficiently plausible explanation?

First, *H* must be causally connected to "*E*, instead of *E**". Obviously, if hemolytic anemia (associated with respiratory infection)

---

[27] Our present task does not require delving deeper into the notion of causation, but for a detailed analysis, see Pearl (2009).

[28] To be sure, identifying causal relations is a delicate business, so our scheme remains simplistic while, at the same time, improving current IBE schemes. Current schemes recognize that cause might be either sufficient or necessary (Walton et al. 2008, p. 185; Wagemans 2016, p. 104), but do not make explicit premises concerning ontological and pragmatic relevance.

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

causally explains the symptoms, then it must, indeed, *cause* dark (instead of normal) urine, high (instead of no) fever, dry (instead of no) cough, etc. This causal connection must be accurate, i.e., supported by the latest medical knowledge.

But this is not enough because *H* might be irrelevant in the case at hand even if it causes "*E*, instead of *E**". Salmon (1971) illustrates this when criticizing Hempel's "covering-law model" of explanation: taking birth control pills avoids pregnancy, but they cannot explain avoiding pregnancy if taken by a male. So, a sufficiently plausible causal explanation requires an accurate and also *ontologically relevant* causal connection: *H* must belong to the causal chain that, in fact, produced "*E*, instead of *E**".

Interestingly, even this does not guarantee that *H* is a sufficiently plausible causal explanation. Ontologically relevant causal chains—i.e., those that produced "*E*, instead of *E**"—consist of many interconnected events. So, how do we pick the explanatorily appropriate cause? Suppose the patient died after being diagnosed with COVID-19. As Lewiński and Abreu (2022) show, explaining the patient's death in terms of COVID-19 faces many challenges. To be sure, sometimes, death cannot be explained by any other cause, but what if there was a preexisting condition, e.g., cancer, contributing to the severe case of COVID-19? Did the patient die "from" COVID-19, or "with" COVID-19? Did cancer cause death? Was it cancer and COVID-19 working together (Lewiński and Abreu 2022, p. 21)? Arguably, selecting the plausible, or appropriate causal explanation might depend on contextual goals, and personal motives (perspectives). For the oncologist, the primary cause of death is most likely cancer. For the government official who wants to raise public awareness of the potentially fatal consequences of contracting COVID-19, the relevant cause of death is COVID-19. A grieving wife, who has been begging her husband to quit cigarettes for nearly thirty years, will find her explanation further up the causal chain: smoking two packs of cigarettes per day. So, a causal connection must be *pragmatically relevant* to the user or the explainee.[29]

[29] Pragmatic relevance is an essential condition of good causal explanation regardless of who the user or the explainee is. In monological contexts, an agent explains things to herself, thereby playing both the explainer's and the explainee's role. In dialogical contexts, one agent assumes the explainer's role and must tailor her

Recognizing pragmatic relevance as a part of the IBE scheme is essential since the early research on explanations and artificial intelligence (Cawsey 1992; Moore 1995). Similarly, contemporary research on the so-called eXplainable Artificial Intelligence (Rohlfing et al. 2021; Nyrup and Robinson 2022) emphasizes that explanations, produced by intelligent systems, must be sensitive to the explainee's background knowledge, goals, needs, and interests. The importance of pragmatic considerations is also recognized by philosophers. Van Fraassen (1980), for instance, cites the following example from Norwood Russell Hanson's (1958, p. 54) *Patterns of Discovery* to elucidate the pragmatic relevance of causal explanations:

> There are as many causes of *x* as there are explanations of *x*. Consider how the cause of death might have been set out by a physician as 'multiple haemorrhage', by the barrister as 'negligence on the part of the driver', by a carriagebuilder as 'a defect in the brakeblock construction', by a civic planner as 'the presence of tall shrubbery at that turning'.

Then, he comments on this passage: "In other words, the salient feature picked out as 'the cause' in that complex process, is salient to a given person because of his orientation, his interests, and various other peculiarities in the way he approaches or comes to know the problem—contextual factors" (van Fraassen 1980, p. 125). Josephson and Josephson (1994, p. 18) reach the same conclusion:

> When we conclude that a finding [*E*] is explained by hypothesis *H*, we say more than just that *H* is a cause of [*E*] in the case at hand. We conclude that among all the vast causal ancestry of [*E*] we will *assign responsibility* [emphasis added] to *H*. Typically, our reasons for focusing on *H* are pragmatic and connected rather directly with goals of production or prevention. We blame the heart attack on the blood clot in the coronary artery or on the high-fat diet, depending on our interests.

---

explanation to the explainee's knowledge, understanding, interests, and needs (see Cawsey 1992; Moore 1995; Walton 2005, 2011) to satisfy the condition of a pragmatic relevance.

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

While such interest relativism can be criticized as woefully subjective, the examples given above point to justifiable, socially constituted rights and obligations of different explanation-seeking and explanation-giving agents, in their recognizable social roles and institutional capacities. So, in terms of Josephson and Josephson (1994), to be a sufficiently plausible explanation, *H* must be a *responsible cause* of "*E*, instead of *E**." And to be a responsible or an *intelligible cause*, *H* must be an accurate, ontologically, and pragmatically relevant cause of "*E*, instead of *E**".[30]

Next, how do we determine if *H* causes "*E*, instead of *E**"? Even if the complete answer to this question were available, we wouldn't be able to present it in one relatively simple argument scheme. In the context of medical diagnostics, however, the key strategy for accepting (or rejecting) some hypothesis comes down to determining to what extent the assumed disease or condition (diagnosis, cause) fits the observed symptoms (explanandum, effect). Setting aside the potential uncertainties regarding the explanandum,[31] the correct diagnosis should neither undergenerate nor overgenerate symptoms. Consider how hemolytic anemia (associated with respiratory infection) gets diagnosed in our example from *casiMedicos*.

Clinical case

> A 32-year-old woman with cerebral palsy from childbirth came to the emergency department after a few days of dark urine associated with an episode of high fever and dry cough. On admission, the CBC showed 16900 leukocytes/mm3 (85% S, 11% L, 4% M), hemoglobin 6.3 g/dL; MCV 109 fl, 360000 platelets/mm3. In the biochemistry LDH 2408; bilirubin 6.8 mg/dl, (unconjugated bilirubin 6.1 mg/dl), normal GOT and

[30] In Lipton's terms, identifying a correct and ontologically relevant cause generates likely explanation, but, to be lovely, explanation must provide understanding, as well. However, what kind of understanding is needed may very well depend on the knowledge, interests, and motives of the explainee, as well as broader contextual goals.

[31] In other words, assuming that there are no *false positives* (inaccurate symptoms that are "detected" due to measurement errors), no *false negatives* (accurate symptoms that are not detected), and that medical examination has been thorough enough.

GPT. The morphological study of blood showed macrocytic anisocytosis with frequent spherocytic forms and polychromatophilia without blasts. The irregular antibody study is positive in the form of panagglutinin, making crossmatching difficult.

Diagnosis

An autoimmune hemolytic anemia would justify the data given: elevated LDH and bilirubinemia due to red cell destruction, polychromatophilia, spherocytosis and anisocytosis because the marrow is working hard to try to compensate for the anemia, which is regenerative. The study of irregular antibodies and the presence of panagglutinin also supports this response, since the binding of an antibody to the hematocyte promotes its lysis and destruction. The girl presents cough and fever, consistent with a respiratory infection.

So, hemolytic anemia (associated with a respiratory infection) gets diagnosed because it explains (most of) the observed symptoms—i.e., does not *undergenarate* (explain only a few among many detected symptoms)—but does not, normally, entail other symptoms that, in the present case, remain undetected (i.e., it does not *overgenerate).* This leads us to the following diagram.

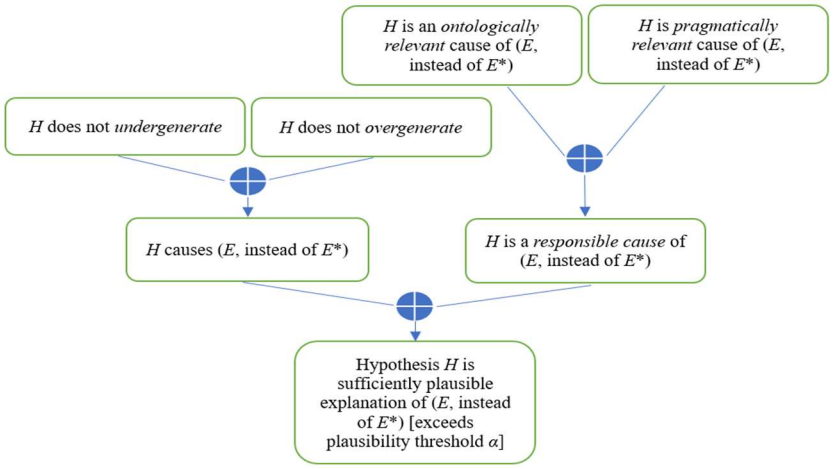

Diagram 7. *IBE: causal explanatory support*

3.6. Rejecting the strongest alternative hypothesis

To complete our IBE scheme, we must tackle one remaining question: How do we reject "the second best", that is, the strongest alternative hypothesis: how do we show that *H**, unlike *H*, is not a sufficiently plausible explanation of "*E*, instead of *E**"?

Previous analysis allows us to give a relatively straightforward answer. Since Diagram 7 identifies positive conditions for accepting *H*, it also shows what can go wrong and represent a reason for rejecting *H**. *H** is rejected if it fails to fulfil some positive condition from Diagram 7: it can undergenerate, overgenerate, or be ontologically or pragmatically irrelevant. Consider how an alternative hypothesis ("The patient has marrow aplasia") gets eliminated in the *casiMedicos* example.

Rejecting alternative diagnosis

> [A] marrow aplasia does not explain the choluria, the elevation of LDH, nor in the study of irregular antibodies is positive in the form of panagglutinin; an aplasia is a marrow failure characterized by a total or partial disappearance of hemopoietic progenitors. In addition, pancytopenia is not observed, which is what would incline us more towards this pathology.

So, basically, marrow aplasia undergenerates (does not explain existing symptoms like choluria, elevated LDH, etc.), as well as overgenerates (would entail pancytopenia that is not detected in the present case), so, plausibly, it does not cause the patient's symptoms. Let us present a hypothesis rejection in structural terms.

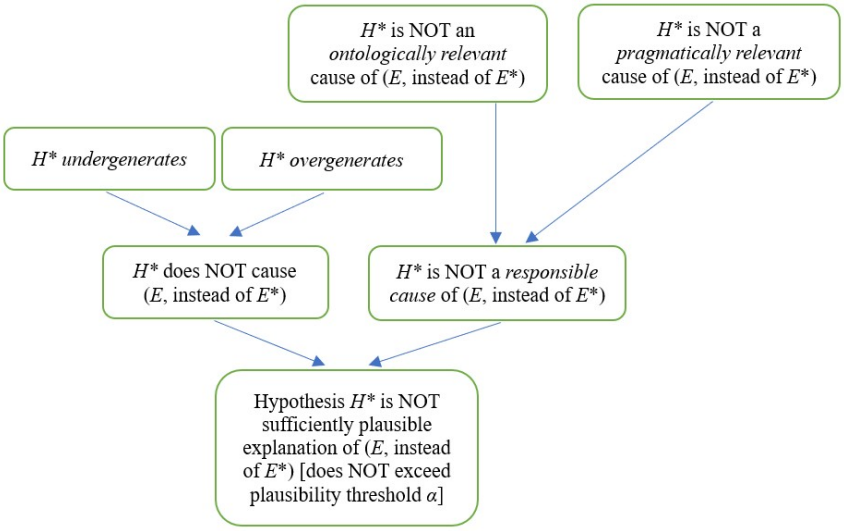

Diagram 8. *IBE: causal hypothesis rejection*

Notice that all (sub)arguments are of the convergent type where "each premise states a separate reason … [and] could support the conclusion without the others" (Govier 2010, p. 38). For instance, even if *H** *were* the cause of "*E*, instead of *E**", rejecting it as an explanation is still justified as long as *H** is not a responsible cause. Or, even if *H** explains all available symptoms (does not undergenerate), rejecting it as a cause is still warranted as long as *H** overgenerates (entails many undetected symptoms).

This concludes our present analysis. We can finally present our improved, extended, and stronger version of the IBE scheme.

4. An extended IBE arguments scheme

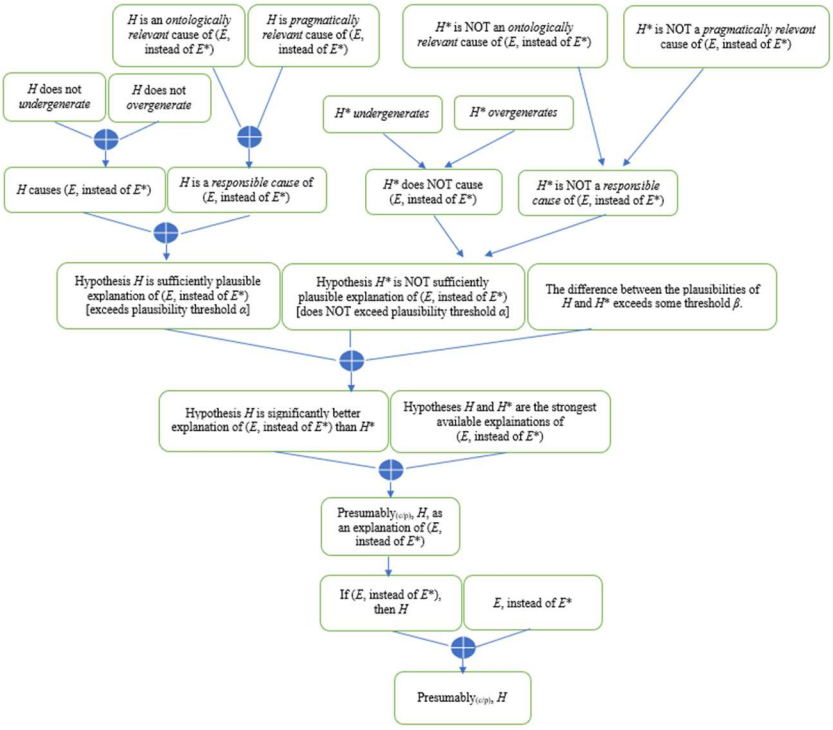

Diagram 9. *IBE: extended scheme (strong formulation)*[32]

5. Applications, implications, and qualifications

But what are the broader implications of our IBE scheme? How does it contribute to the ongoing research in philosophy, argumentation theory, and artificial intelligence?

The proposed IBE scheme contributes to argumentation theory in several ways. Most obviously, it avoids several shortcomings associated with the previous IBE schemes: it represents IBE's premises and conclusions more accurately (by more careful use of modifiers), reduces structural vagueness (by using diagrams and referring to standard argument patterns), and introduces several premises that were not incorporated in previous schemes (despite being recognized

[32] Full-sized version available in the appendix on p. 682.

in the literature). Second, it advances the normative study of reasoning by going beyond the well-known "critical questions approach." In argumentation theory in general and IBE scholarship in particular, it is quite common to provide a list of standard critical questions associated with a particular scheme. Some piece of reasoning (that instantiates the scheme) sufficiently supports the conclusion only if such questions are answered appropriately (Walton 2001; Walton et al. 2008; Yu and Zenker 2018; Olmos 2021). For instance, Walton identifies four critical questions associated with the abductive argumentation scheme. We mention only two:

> CQ1: How satisfactory is [*H*] itself as an explanation of [*E*], apart from the alternative explanations available so far in the dialogue?
> CQ2: How much better an explanation is [*H*] than the alternative explanations available so far in the dialogue? (Walton 2001, p. 162)

Although normative assessment based on critical questions is plausible, pedagogically useful, and dialectically appropriate, it also fails to specify a logical relationship between normative requirements. For instance, Walton's CQ1 recognizes the *non-comparative*, and CQ2 a *comparative requirement* of the explanation's quality, but his question-based approach does not elucidate either the logical relationship between the two requirements or between such requirements and the conclusion they attempt to support. Once we represent normative requirements (that underlie critical questions) as additional premises, we become able to provide even clearer instructions for the evaluation of IBE. For instance, according to our IBE scheme, a non-comparative premise "*H* is sufficiently plausible explanation of '*E* instead of *E**'" partly justifies a comparative premise "*H* is a significantly better explanation of '*E*, instead of *E**' than *H**." Accordingly, the evaluator should pay more attention to a more fundamental, non-comparative explanatory strength. Since it elucidates normative hierarchy and provides even more practical instructions for assessing IBE, our present analysis favors a premise-based over a question-based approach. Admittedly, however, each of the scheme's (sub-)premises can be an object of critical questioning, as

 Informal Logic, Vol. 44, No. 4 (2024), pp. 636–682.

delineated in the argumentation scheme tradition. We hope to give this issue its due attention in future work.[33]

Third, to some extent, our scheme bridges the gap between studying inference (argument, explanation) as a *product* and studying it as a *process*, a distinction well-known in philosophy and argumentation theory (see O'Keefe 1977; Wenzel 1992; Grimm 2010). To be sure, what we develop here is IBE's inferential scheme, i.e., the final product: a strong or successful IBE is such, precisely because the elements of the complex scheme are fully considered and defended (see Sect. 4). However, the scheme also allows us, at least to some extent, to understand IBE as a process. The process, of course, is extended in time and starts from the initial "brainstorming" of a pool of (however weak) hypotheses and concludes in a justification of the presumptive truthfulness of "the best explanation", that is, the explanatory hypothesis that is best corroborated in the verification process. In other words, along the temporal axis, the process moves from the weak entitlements and obligations of a "mere hypothesis" to the strong ones of a successfully verified, "best explanation". This explanation, however "best" it is for now, can always be revised or even entirely controverted as new evidence or inferences flow in; yet it carries the presumptive status of the best truth candidate unless and until it is revised or controverted. Although our scheme is not temporal, it allows us to "follow" IBE as a process by keeping track of modal qualifiers. In short, *H* starts as a *plausible hypothesis* (yet to be tested), becomes the best, *presumed explanation*, and, finally, on these explanatory grounds, gets *presumed as a true proposition*.

Fourth, the extended scheme enables us to closely study the interplay between explaining and justifying (or explanations and arguments), a subject relevant to both philosophers and argumentation scholars. Notice that we treat explanandum as a basic premise of IBE, a proposition that justifies the conclusion, the (presumptive) explanans. By contrast, Hempel (1966) treats an explanandum as a conclusion that is being explained by the set of premises (initial conditions and natural laws). Our approach has consequences for Hempel-derived textbook discussions of the argumentation-

[33] We thank the anonymous reviewer for pushing us to clarify our stance on the "critical questions approach."

explanation distinction in argumentation theory (cf. Dufour 2017; McKeon 2013), where for argumentation conclusion is treated as controversial and thus an object of a dispute, whereas in explanations conclusion is uncontroversial, as the mutually observed fact to be explained. On our approach, IBE is, somewhat paradoxically, argumentative through and through, as it's the "bestness" of the conclusion-qua-explanans that is at issue here, rather than the justification of the explanandum. In short, the inference from "*E* instead of *E**" (explanandum) to "Presumably, *H*" (ultimate conclusion) is justified by *H*'s explanatory "bestness," and the "bestness" is justified by various comparative and non-comparative premises of a complex *argumentation* scheme.

Fifth, our scheme might contribute to the philosophical discussion about realism and anti-realism, although modestly and indirectly. In the philosophy of science, realists contend that good scientific explanations attempt to give us knowledge (accurate descriptions) of the world, i.e., that acceptable scientific explanations must consist of true beliefs. By contrast, anti-realists argue that good explanations are (also) associated with various pragmatic ends, do not attempt to describe mind-independent reality, and that, consequently, the explanations' acceptability requires neither believing that the explanations are true, nor that the entities they assume, in fact, exist. Our scheme is general enough to capture both realistic (e.g., "*H* is an ontologically relevant cause") and anti-realistic requirements (e.g., "*H* is a pragmatically relevant cause") and approximate their logical relationships.

Also, the idea that IBE's conclusion is aimed at a correct description of the explanandum need not entail any strong realist commitments about the ontological status of "the unobservable." The "truth" of the explanans (cf. Harman) can have strong realist reading (entities that explain actually exist), but a range of more modest interpretations—still avoiding deflationary nominalism—is available. One of them is van Fraassen's (1980) "constructive empiricism" content with the notion that "*the acceptance of a theory involves as belief only that it is empirically adequate*" (1980, p. 12, italics in the original). In many quotidian and also medical cases we discussed here, the "empirical adequacy" of explanations bottoms out in some actually existing phenomena: the noises in the wall are explained by the

fact "that a mouse has come to live with me" (van Fraassen 1980, p. 20); the symptoms are explained by the fact that a patient has contracted virus X. But in more sophisticated theoretical cases discussed by the philosophers of science interested in IBE,[34] the inference from best explanation to factual existence is not so straightforward and, therefore, remains highly controversial. Our approach can be comfortably non-committal on such issues because, first, we're dealing with the *metaphysically* easy cases here and, second, because our goal is to understand the argumentative and pragmatic nature of IBE's conclusion, not its metaphysical status. As such, it remains open to and consistent with a variety of views.[35]

Sixth, systematizing IBE's normative (evaluative) conditions is also relevant for understanding abductive reasoning. The relationship between IBE and abduction is philosophically controversial since scholars associate them with different stages of scientific inquiry and emphasize differences between the context of discovery and the context of justification, that is, between the generation and evaluation of hypotheses (see Hintikka 1998; Minnameier 2004; Campos 2009; Mcauliffe 2015; Yu and Zenker 2018; Urbański and Klawiter 2018). Intuitively, however, generation (associated with abduction) and evaluation (associated with IBE) are inextricably intertwined: while the creative process of producing a (plausible) hypothesis always involves evaluation—whereby unpromising hypotheses are quickly discarded, or not even imagined—the evaluative process of comparing-and-contrasting a hypothesis against other candidate hypotheses both relies on and possibly inspires further generation (see Magnani 2001). So, from the structural viewpoint, conditions that, in the case of IBE, justify H as the best explanation (that should be presumed at the final stage of inquiry) may also, in the case of abduction, justify $\{H, H_1, H_2, H_3\}$ as the best *set of potential explanations* (that should be considered at the initial stage of inquiry and tested further on). Even if abduction and IBE are not identical types of reasoning (as they are applied at different stages of the inquiry), they still seem structurally similar, so studying how, ultimately, the

[34] For instance, do atoms, black holes, and attractive forces actually exist, or are rather theoretical constructs capable of explaining the observable phenomena?

[35] We thank the anonymous reviewer for pushing us to clarify the relationship between our IBE scheme and the realist approach.

best hypothesis is selected might help us understand how, initially, plausible hypotheses are generated.[36]

Finally, we believe that a detailed structural account of IBE may contribute to the development of the so-called eXplainable Artificial Intelligence (XAI). XAI is an emerging field in AI that seeks to develop methods that would make AI systems transparent and interpretable by clarifying their goals, training, databases, limitations, and inferences. As Mittelstadt, Russell and Wachter (2018) remark, everyone agrees that AI systems should explain their outputs—e.g., medical diagnostic AI systems should be able to explain why and how their diagnoses are inferred from the input data—without having a clear idea of what, exactly, constitutes a good explanation.

> [O]ne of the most striking aspects of research into explainable AI (xAI) is how many different people, be they lawyers, regulators, machine learning specialists, philosophers, or futurologists, are all prepared to agree on the importance of explainable AI. However, very few stop to check what they are agreeing to, and to find out what explainable AI means to other people involved in the discussion. (Mittelstadt, Russell and Wachter 2018)

Hopefully, IBE's structural account might at least sometimes offer guidelines to XAI researchers and direct AI systems toward human-like explanations based on causal (instead of statistical) considerations.

> If we want to design, and implement intelligent agents that are truly capable of providing explanations to people, then it is fair

[36] Olmos (2021) distinguishes between abduction and IBE not in terms of internal reasoning processes, but rather as forms of externalized arguments. Abduction is a form of *argumentation scheme* with its unique justificatory *warrant* (namely: that the generated hypothesis could explain the data), while IBE is a "communicative processes of selecting, assessing, and weighing alternative explanatory hypotheses, purporting the use of varied *argumentative*, *counterargumentative*, and *metaargumentative schemes* and *structures*" (2021, p. 136). Even on this account, most useful in better understanding the argumentative dynamics of abduction and IBE, the structural similarity between them holds: abductions would be characteristically nested within IBE activities, while the latter would be inescapably, even if only partly, constituted by the former.

> to say that models of how humans explain decisions and behaviour to each other are a good way to start analysing the problem. (Miller 2019, p. 2)

> To reach a level of *explainable medicine* we need causability … [that] encompasses measurements for the quality of explanations [Causibility is] the extent to which an explanation of a statement to a human expert achieves a specified level of causal understanding with effectiveness, efficiency and satisfaction in a specified context of use. (Holzinger et al. 2019, pp. 1-3)

Our IBE scheme identifies typical premises that the explainee might need in order to understand the explanandum or to test the initial explanation provided by the AI system. Since such premises can also be understood as responses to explainee's (critical) questions, XAI researchers might directly profit from our systematic account of IBE. For a short literature review discussing ways in which argument schemes can be relevant to different lines of research in AI, see Macagno (2021).

6. Conclusion

We constructed a new IBE scheme that avoids several problems associated with present schemes, such as inadequate representations of propositions, structural vagueness, and unwarranted simplicity.

First, the proposed scheme avoids some problems of representation by a more precise treatment of pragmatic (epistemic) modifiers ("hypothetically," "plausibly," and "presumably"), and acknowledging the contrastive nature of explanandum. Second, it eliminates structural vagueness by using diagrams and clarifying the logical support between premises and conclusions (linked, convergent, and linear support). Finally, the new IBE scheme reduces unwarranted simplicity by introducing additional premises essential for inferring justified conclusions (e.g., premises related to the higher standards of proof, ontological and pragmatic relevance, under- and overgeneration, etc.).

By proposing a comprehensive checklist of normative concerns essential for performing and evaluating IBE, the present scheme

contributes to understanding a ubiquitous type of human reasoning (and also allows for computational applications, in a natural connection to the emerging field of XAI). In future analysis, we should improve our simplistic account of causal explanation and generalize the IBE scheme, so that it equally applies to non-causal (e.g., statistical) instances of IBE.

Acknowledgements: We would like to thank audiences at the 13th Oficina de Filosofia Analítica (Lisbon, Portugal, November 2023), 13th Conference of the Ontario Society for the Study of Argumentation (OSSA) on "Argumentation and Changing Minds" (University of Windsor, Ontario, Canada, May 2024) and the 4th Argumentation & Language Conference (ARGAGE) on "Pragmatics and Argumentation" (University of Fribourg, Switzerland, June 2024), as well as two anonymous reviewers for *Informal Logic*, for their most useful comments and suggestions. This work has been supported by the Portuguese Foundation for Science and Technology (FCT), CHIST-ERA Programme via the project ANTIDOTE: Argumentation-Driven Explainable Artificial Intelligence for Digital Medicine (CHIST-ERA/0002/2019), and the project Antipsychologistic conceptions of logic and their reception in Croatian philosophy (APsiH) at the Institute of Philosophy, Zagreb (reviewed by the Ministry of Science and Education of the Republic of Croatia and financed through the National Recovery and Resilience Plan by the European Union—NextGenerationEU).

References

Bodlović, P. 2021. On the differences between practical and cognitive presumptions. *Argumentation* 35(2): 287-320. https://doi.org/10.1007/s10503-020-09536-w

Bodlović, P. 2022. On the strength of presumptions. *Pragmatics & Cognition* 29(1): 82-110. https://doi.org/10.1075/pc.21017.bod

Bodlović, P. and K. Kudlek. 2024. Knowledge versus understanding: what drives moral progress? *Ethical Theory and Moral Practice*, online first. https://doi.org/10.1007/s10677-024-10465-w

Campos, D. G. 2011. On the distinction between Peirce's abduction and Lipton's inference to the best explanation. *Synthese* 180: 419-442. https://doi.org/10.1007/s11229-009-9709-3

Cawsey, A. 1992. *Explanation and interaction. The computer generation of explanatory dialogues*. The MIT Press: Cambridge, Massachusetts.

Douven, I. 2021. Abduction. In *The Stanford encyclopedia of philosophy* (summer 2021 edition), ed. Edward N. Zalta. URL accessed 15 September 2022: https://plato.stanford.edu/archives/sum2021/entries/abduction/ .

Dragulinescu, S. 2016. Inference to the best explanation and mechanisms in medicine. *Theoretical Medicine and Bioethics* 37: 211–232. https://doi.org/10.1007/s11017-016-9365-9

Dretske, F. I. 1972. Contrastive statements. *The Philosophical Review* 81(4): 411-437.

Dufour, M. 2017. Argument or explanation: Who is to decide? *Informal Logic* 37(1): 23-41. https://doi.org/10.22329/il.v37i1.4523

van Eemeren, F.H. and R. Grootendorst. 2004. *A systematic theory of argumentation. The pragma-dialectical approach*. Cambridge: Cambridge University Press.

Elgin, C. 2007. Understanding and the facts. *Philosophical Studies* 132(1): 33-42. https://doi.org/10.1007/s11098-006-9054-z

van Fraassen, B.C. 1980. *The Scientific Image*. Oxford: Clarendon Press.

van Fraassen, B.C. 1989. *Laws and symmetry*. Oxford: Clarendon Press.

Freeman, J.B. 2005. *Acceptable premises*. Cambridge: Cambridge University Press.

Gaszczyk, G. 2023. Helping others to understand: A normative account of the speech act of explanation. *Topoi* 42: 385-396. https://doi.org/10.1007/s11245-022-09878-y

Godden, D. 2017. Presumption as a modal qualifier: Presumption, inference, and managing epistemic risk. *Argumentation* 31(3): 485-511. https://doi.org/10.1007/s10503-017-9422-1

Godden, D. and D. Walton. 2007. A theory of presumption for everyday argumentation. *Pragmatics and Cognition* 15 (2): 313-346.
Govier, T. 2010. *A practical study of argument* (7th edition). Wadsworth: Cengage Learning.
Grimm, S. 2010. The goal of explanation. *Studies in History and Philosophy of Science* 41: 337-344. https://doi.org/10.1016/j.shpsa.2010.10.006
Hanson, N. R. 1958. *Patterns of Discovery*. Cambridge: Cambridge University Press.
Harman, G. H. 1965. The inference to the best explanation. *The Philosophical Review* 74(1): 88–95. https://doi.org/10.2307/2183532
Hempel, C. 1966. *The philosophy of natural science*, Englewood Cliffs: Prentice Hall.
Hills, A. 2016. Understanding why. *Noûs* 49(2): 661-688. https://doi.org/10.1111/nous.12092
Hintikka, J. 1998. What is abduction? The fundamental problem of contemporary epistemology. *Transactions of the Charles S. Peirce Society* 34(3): 503-533.
Holzinger, A., Langs, G., Denk, H., Zatloukal, K. and H. Müller. 2019. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov*, 9, 4, e1312. doi: 10.1002/widm.1312. Epub 2019 Apr 2. PMID: 32089788; PMCID: PMC7017860.
Josephson, J. R. and S. G. Josephson. 1994. *Abductive Inference: computation, philosophy, technology*. New York: Cambridge University Press.
Lewiński, M. 2017. Argumentation theory without presumptions. *Argumentation* 31(3): 591–613. https://doi.org/10.1007/s10503-017-9421-2
Lewiński, M. and P. Abreu. 2022. Arguing about "COVID": Metalinguistic arguments on what counts as a "COVID-19 Death." In *The pandemic of argumentation*, eds. S. Oswald, M. Lewiński, S. Greco, S. Villata, 17-43. Springer. https://doi.org/10.1007/978-3-030-91017-4_2
Lewis, D. 1986. Causal explanation. In his *Philosophical papers* (Vol. 2). New York: Oxford University Press.
Lipton, P. 2004. *Inference to the best explanation* (2nd edition). Routledge.
Macagno, F. 2021. Argumentation schemes in AI: A literature review. Introduction to the special issue. *Argument and Computation* 12(3): 287-302.
Macagno, F. and D. Walton. 2012. Presumptions in legal argumentation. *Ratio Juris* 25(3): 271-300. https://doi.org/10.1111/j.1467-9337.2012.00514.x

© Petar Bodlović and Marcin Lewiński. *Informal Logic,* Vol. 44, No. 4 (2024), pp. 636–682.

Magnani, L. 2001. *Abduction, reason and science*. *Processes of discovery and explanation*. Dordrecht: Kluwer.

Marcum, J. A. 2008. *Humanizing Modern Medicine. An Introductory Philosophy of Medicine*. Springer Science + Business Media B.V.

Mcauliffe, W. H. B. 2015. How did abduction get confused with inference to the best explanation? *Transactions of the Charles S. Peirce Society* 51(3): 300-319.

McKeon, M. W. 2013. On the rationale for distinguishing arguments from explanations. *Argumentation* 27: 283–303. https://doi.org/10.1007/s10503-012-9288-1

Miller, T. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267: 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Minnameier, G. 2004. Peirce-suit of truth: Why inference to the best explanation and abduction ought not to be Confused. *Erkenntnis* 60 (1): 75-105. https://doi.org/10.1023/B:ERKE.0000005162.52052.7f

Mittelstadt, B., Russell, C. and S. Wachter. 2019. Explaining explanations in AI. arXiv:1811.01439 [cs.AI]: 1-10.

Moore, D. J. 1995. *Participating in explanatory dialogues. Interpreting and responding to questions in context.* Cambridge, Massachusetts: The MIT Press.

Nyrup, R. and D. Robinson. 2022. Explanatory pragmatism: a context-sensitive framework for explainable medical AI. *Ethics and Information Technology* 24: 13. https://doi.org/10.1007/s10676-022-09632-3.

O'Keefe, D. J. 1977. Two concepts of argument. *Journal of the American Forensic Association* 13(3): 121-128.

Olmos, P. 2021. Metaphilosophy and argument: The case of the justification of abduction. *Informal Logic* 41(2): 131–164. https://doi.org/10.22329/il.v41i2.6249

Pearl, J. 2009. *Causality: Models, reasoning, and inference* (2nd edition). New York: Cambridge University Press

Peirce, C. S. 1903. Harvard Lectures on Pragmatism. *In The collected papers of Charles Sanders Peirce* (electronic edition), Vol. 5, eds. P. Weiss, C. Hartshorne, and A.W. Burks. Cambridge, MA: Harvard University Press.

Peirce, C. S. 1994. *The collected papers of Charles Sanders Peirce* (electronic edition), Vol. 1–8), eds. P. Weiss, C. Hartshorne, and A.W. Burks. Cambridge, MA: Harvard University Press.

Prakken, H. and G. Sartor. 2009. A logical analysis of burdens of proof. In *Legal evidence and proof: statistics, stories, logic*, eds. H. Kaptein, H.

Prakken, B. Verheij, 223-253. Farnham: Ashgate Publishing (Applied Legal Philosophy Series).
de Regt, H. 2015. Scientific understanding: truth or dare? *Synthese* 192(12): 3781–3797. https://doi.org/10.1007/s11229-014-0538-7
Rescher, N. 1976. *Plausible reasoning*. Assen/Amsterdam: Van Gorcum.
Rescher, N. 2006. *Presumption and the practices of tentative cognition*. Cambridge: Cambridge University Press.
Rohlfing K. J. et al. 2021. Explanation as a social practice: toward a conceptual framework for the social design of ai systems. *IEEE Transactions on cognitive and developmental systems* 13(3): 717-728. doi: 10.1109/TCDS.2020.3044366
Salmon, W. 1971. Statistical explanation. In *Statistical explanation and statistical relevance*, ed. W. Salmon, 29-87. Pittsburgh, PA: University of Pittsburgh Press.
Ullmann-Margalit, E. 1983. On presumption. *Journal of Philosophy* 80(3): 143-163.
Urbański, M. and A. Klawiter. 2018. Abduction: Some conceptual issues. *Logic and Logical Philosophy* 27(4): 583-597.
Wagemans, J.H.M. 2016a. Criteria for deciding what is the 'best' scientific explanation. In *Argumentation and reasoned action*: *proceedings of the 1st European conference on argumentation* (Lisbon, 2015), Vol. II, eds. D. Mohammed, M. Lewiński, 43-53. London: College Publications.
Wagemans, J.H.M. 2016b. Argumentative patterns for justifying scientific explanations. *Argumentation* 30: 97-108. https://doi.org/10.1007/s10503-015-9374-2
Walton, D. 2001. Abductive, presumptive and plausible arguments. *Informal Logic* 21(2): 141-169. https://doi.org/10.22329/il.v21i2.2241
Walton, D. 2005. *Abductive reasoning*. Tuscaloosa, Alabama: The University of Alabama Press.
Walton, D. 2006. *Character evidence. An abductive theory*. Dordrecht: Springer.
Walton, D. 2011. A dialogue system specification for explanation. *Synthese* 182: 349-374. https://doi.org/10.1007/s11229-010-9745-z
Walton, D., C. Reed and F. Macagno. 2008. *Argumentation schemes*. Cambridge: Cambridge University Press.
Wenzel, J. W. 1992. Perspectives on argument. In *Readings in Argumentation*, eds. W. L. Benoit, D. Hample and P. J. Benoit, 121–143. Berlin / New York: Foris.
Wilholt, T. 2009. Bias and values in scientific research. *Studies in History and Philosophy of Science Part A* 40(1): 92-101. https://doi.org/10.1016/j.shpsa.2008.12.005

Witek, M. 2021. Illocution and accommodation in the functioning of presumptions. *Synthese* 198(7): 6207–6244.

Woods, J., A. Irvine and D. Walton. 2003. *Argument: critical thinking, logic, and the fallacies* (2nd Edition). Toronto: PEARSON, Prentice Hall.

Yu, S. and F. Zenker. 2018. Peirce knew why abduction isn't IBE—A scheme and critical questions for abductive Argument. *Argumentation* 32: 569-587. https://doi.org/10.1007/s10503-017-9443-9

Yu, S. and F. Zenker. 2022. Identifying linked and convergent argument structures: A problem unsolved. *Informal Logic* 42 (2): 363-387. https://doi.org/10.22329/il.v42i1.7133

Appendix

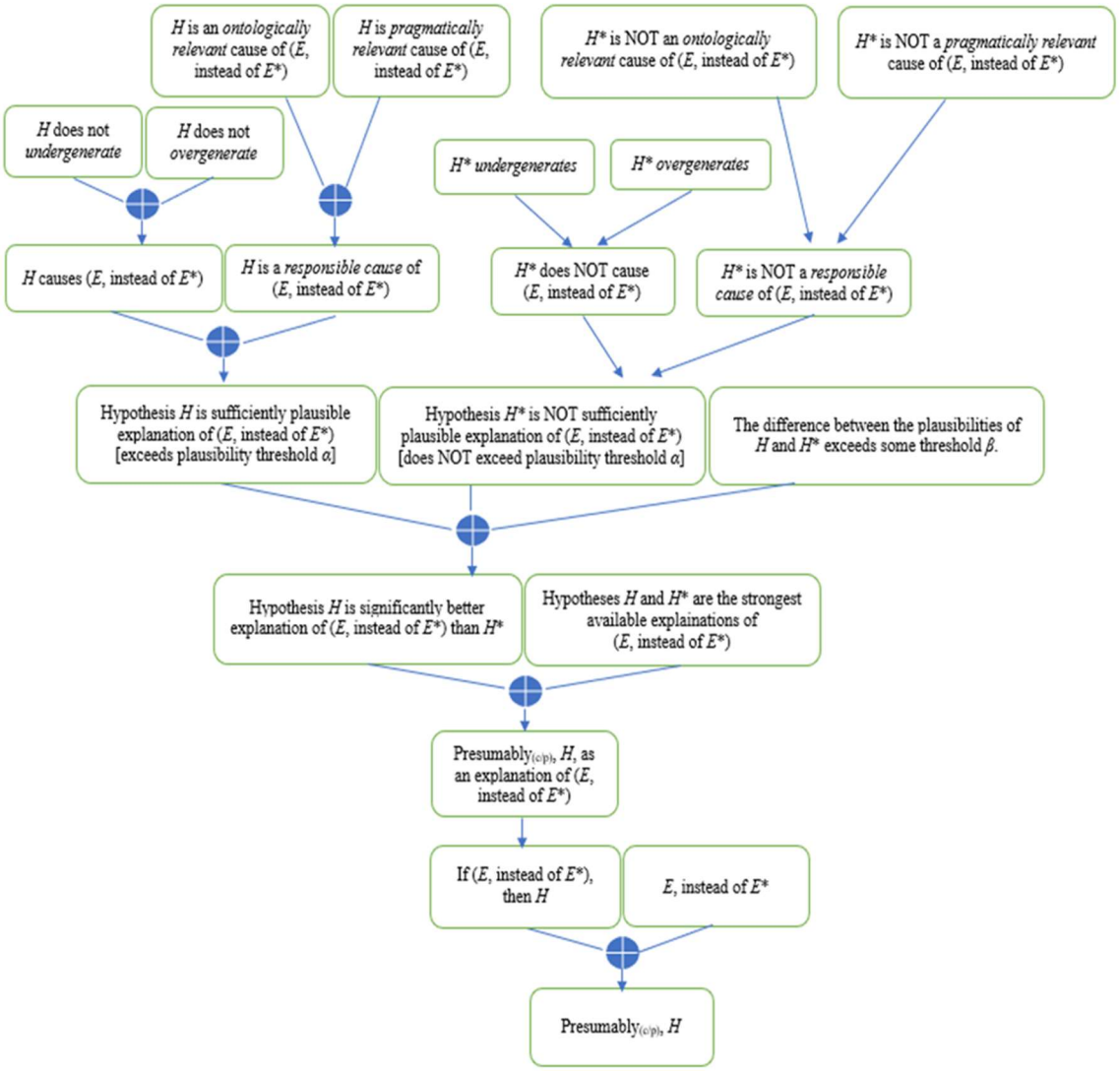

Diagram 10. Version of. *IBE: extended scheme (strong formulation) (Enlarged)*