Journal of Applied Non-Classical Logics
Vol. 00, No. 00, Month 201X, 1–46

RESEARCH ARTICLE

A Logic for Best Explanations

Author's Manscript. Please cite published version at DOI: https://www.tandfonline.com/doi/full/10.1080/11663081.2019.1591108.

Jared Millson[a*] and Christian Straßer[b]

[a]*Agnes Scott College, Decatur, GA, USA*; [b]*Ruhr-Universität Bochum, Germany*

(*v1.0 released September 2012*)

There are two obstacles confronting efforts to formalize qualitative accounts of inference to the best explanation (IBE). The first is that extant accounts of IBE in philosophy of science remain painfully imprecise and often rely on contextual criteria to evaluate explanations. The second is that our best explanations exhibit unusual logical properties, such as contradiction-intolerance and irreflexivity. This paper aims to surmount these challenges by utilizing a new, more precise theory that treats explanations as expressions that codify defeasible inferences. To formalize this account, we provide a sequent calculus in which IBE serves as an elimination rule for a connective that exhibits many of the properties associated with the behavior of the English expression "That. . . best explains why. . . ". We first construct a calculus that encodes these properties at the level of the turnstile, i.e. as a metalinguistic expression for classes of defeasible consequence relations. We then show how this calculus can be conservatively extended over a language that contains a *best-explains-why* operator.

Keywords: abductive inference; inference to the best explanation; scientific explanation; defeasible reasoning; sequent calculus

1. Introduction

After suffering decades of neglect, abductive reasoning has at last become a respectable philosophical preoccupation. Among epistemologists and philosophers of science, this renewal of interest has focused on the form of abduction known as *inference to the best explanation* (IBE).[1] Most work in this area falls into one of two camps: qualitative approaches that treat IBE as a non-deductive species of logical inference, and quantitative approaches that view IBE as a form of probabilistic reasoning. While quantitative treatments have found success formalizing IBE (Douven, 1999, 2002; Glass, 2007; Schupbach & Sprenger, 2011; Weisberg, 2009), qualitative treatments remain largely informal (Harman, 1965; Lipton, 2004; Psillos, 2000, 2002). Similarly, in the fields of logic and

*Email: jmillson@agnesscott.edu

[1]There does not appear to be a consensus on whether abduction and IBE are distinct reasoning patterns. (For example, see Campos (2011)). In this paper, we will treat 'abduction' as a generic term covering a panoply of non-deductive inference patterns, not all of which include explanatory premises. (For a conception of abduction that does not restrict the notion to explanatory reasoning, see Gabbay and Woods (2006).) 'IBE' refers to just one (or one class) of these patterns, namely, that which is typically the subject-matter of debates in philosophy of science (Harman, 1965; Lipton, 2004; Psillos, 2000, 2002) and which is identified with the schema: A *best explains* $B, B / \therefore A$. Here is another, hopefully helpful specification: our target inference corresponds to 'consistent, explanatory, preferential abduction' in Aliseda's (2006) taxonomy of abduction styles.

computer science, where formal, qualitative accounts of abduction proliferate, formalizations of IBE are quite rare—a fact all the more surprising given the diversity that that literature boasts in both perspectives (structural, adaptive, tableau-based, belief revision, etc.) and targets—i.e. *types* of abduction (creative vs. selective, consistent vs. anomalous, factual vs. nomic, etc.).[1] This paper aims to fill this lacuna by formalizing a qualitative treatment of IBE.

Why have there been so few attempts to logically formalize IBE? One reason has to do with the fact that what constitutes a *best explanation* remains far from settled. Worse yet, extant accounts of this concept are nowhere near the level of precision needed to meet logicians even half-way. Finally, many philosophers of science believe that our best explanations are determined, in part, by contextual considerations (Day & Kincaid, 1994; Lipton, 2004). Consequently, a formalization of IBE will likely have to reserve a parameter for extra-logical criteria. There have been some attempts to do so by incorporating structured preference relations (Boutilier & Becher, 1995; Lobo & Uzcátegui, 1997; Mayer & Pirri, 1996; Pino-Pérez & Uzcátegui, 1999, 2000, 2003), but such an approach treats the evaluation of explanantia as *given* rather than as a product of reasoning. At the very least, this makes the distinctively evaluative character of IBE rather uninteresting from a logical perspective.

Disagreement among philosophers of science, is only partly to blame for logicians' neglect of IBE. For even the most basic, least controversial properties of explanatory relations present daunting challenges to formalization. Explanations appear to be contradiction-intolerant, (possibly) non-transitive, (definitely) non-monotonic, and even irreflexive. Some of these properties, such as non-transitivity and non-monotonicity are captured by well-established logics, but others, such as irreflexivity, run counter to the most liberal conceptions of logical consequence.[2] As a result, logic-based treatments of non-evaluative, 'plain' abduction shy away from directly encoding explanatory relations in their object-language.[3] Instead, they simply use material implication as a stand-in for the explanatory connective and then impose constraints on the application of 'backwards Modus Ponens' in order to block reflexive, contradictory, or overly-informative explanations.[4] Since the traditional gloss of IBE—i.e. A *best explains* B, B; so, A—gives explanatory vocabulary pride of place, a logical treatment of IBE is likely to face all of

[1] Taxonomies of abductive reasoning can be found in Aliseda (2006); Magnani (2001); Schurz (2008). For excellent examples of some of the different formal approaches to abduction, see Magnani, Carnielli, and Pizzi (2013).

these challenges, if not more.

There are thus two sets of obstacles confronting the logical formalization of IBE. The first lies in the philosophy of science: what characterizes our best explanations? The second belongs to logic proper: how do we perspicuously codify explanatory relations and their evaluation? While this paper offers a formal treatment of IBE and, in doing so, aims to meet head-on the latter set of challenges, it does so as part of a broader effort to provide a compelling and precise qualitative account of best explanations. We call this account *the defeasible inference model of explanation* (DIME).

Motivating this project is the idea that explanatory relations should neither be reduced to material implication nor treated as primitives. Proponents of IBE in philosophy of science (a.k.a explanationists) have long held that a prior grip on explanation can underwrite good ampliative inferences. On our view, this gets things fundamentally backwards: *contra* explanationism, inquirers first engage in inference-making, and only late in the game do they baptize some of these inferences as explanations. The priority of inference over explanation captivated those earlier theories of explanation—e.g. Friedman (1974); Hempel (1965); Kitcher (1989); Schurz and Lambert (1994)—that belong to a family of views we call Explanation-as-Inference (EAI). Unfortunately, these theories tended to assimilate the explanatory relation to deductive consequence. What sets our approach apart is the fact that we treat explanatory claims as expressing or making explicit certain *non-deductive* inferences.

Our proposal is to treat the explanatory connective in IBE as an object-language expression for what, in the meta-language, is a rule of defeasible inference. Roughly put, the *best-explains-why* operator codifies a defeasible consequence relation just as material implication codifies the classical consequence relation. Since both abduction and explanatory relations are recognized to be defeasible, such a model provides a natural starting point for their joint formalization. By specifying conditions of inferential defeat, we are able to circumscribe classes of consequence relations that exhibit the properties associated with our best explanations. Rather than appeal to a pre-established preference ranking, we use a model of defeasible inference that has two parameters for extra-logical information: one for the set of situations in which an inference fails and another for information on the actual situation in which the inference is made, yet distinct from that contained in the premises. By formalizing operations on these two parameters,

[2]Many of these exotic properties have been investigated in recent attempts to formalize metaphysical explanations in what has come to be known as the *logic of grounding* (Fine, 2012; Litland, 2016; Schnieder, 2011). In particular, logics have been proposed for full strict immediate ground and similar relations that exhibit irreflexivity, asymmetry, non-transitivity, and (in some sense) minimality (Poggiolesi, 2018). While we remain suspicious regarding the extent to which this relation is genuinely *explanatory*, we have been inspired by the innovative work done by logicians in this area.

[3]There are two notable exceptions to this trend. The first is a logic of explanatory relations based on structural rules developed by Pino-Pérez and Uzcátegui (1999, 2000, 2003). This logic contains a connective for the explanatory relation that is closed under several versions of cut. Together with a rule of cautious monotonicity, these cut rules encode a simplicity criteria that induces a linear pre-order over explanations. While this system yields many desirable results, the rules employ relations, such as $\nvdash$, that are not well-behaved. Moreover, some of the properties attributed to explanatory relations, such as reflexivity and cautious monotonicity, we find implausible. The second noteworthy work is that of Mathieu Beirlaen and Atocha Aliseda (2014); it is perhaps the most successful attempt to do justice to the explanatory character of abductive premises. By using the conditional from Brian Chellas's ultra-weak **CR** to codify the explanatory relation, Beirlaen and Aliseda are able to block the validity of *idempotence*, *ex falso quod libet*, *strengthening the antecedent*, *modus ponens*, *contraposition*, and *hypothetical syllogism*—to name just a few. The result is an adaptive logic for (plain) abduction that operates on conditionals that behave in many ways like explanatory expressions do in English. There is, however, no introduction rule for their explanatory conditional—i.e. there is no sense in which one might reason *to* an explanatory proposition. The system offered in the present work is an attempt, *inter alia*, to capture this missing feature. Insofar as it gives a richer formalization of abductive premises, the work of adaptive logicians utilizing modal languages is also noteworthy (Meheus, 2005; Meheus & Batens, 2006; Meheus, Verhoeven, Van Dyck, & Provijn, 2002).

[4]A canonical example of such an approach is Flach (1996a, 1996b).

we provide a uniform proof theory that captures both the structure of IBE as well as the least controversial properties of the *best-explains-why* connective.

Among the novel contributions of DIME is the property of *sturdiness*, which it imputes to our best explanations. Many have claimed that explanations must be *stable* in the sense that the purported relationship between explanans and explanandum must persist under a range of different conditions (Hempel, 1965; Lange, 2009; Mitchell, 2003; Skyrms, 1980; Woodward, 2003). Sturdiness is a particular way of thinking about the stability of explanatory claims when the latter are treated as commitments to defeasible inferences. In slogan form, an inference is sturdy just in case it succeeds where its competitors fail. In section 3, we unpack this slogan. By formalizing what we call the "sturdiness test", our calculus captures this crucial property and uses it to realize other important features of explanations, such as minimality.

The formal work presented here assumes that DIME is the correct account of best explanations. As the name suggests, DIME belongs to the EAI family. Recent work on this theory has sought to clear the ground for a renewal of the EAI view (Khalifa, Millson, & Risjord, 2018b), to defend the model from criticisms that scuttled previous versions of EAI (Khalifa, Millson, & Risjord, 2018a), and to situate our approach within a broader theory of logical vocabulary (Millson, Khalifa, & Risjord, 2018). In this paper, we have a narrower target in view, namely, to present the details of our formalization of IBE and provide a worked-example from the literature, to demonstrate that the explanatory connective encodes (a class of) consequence relations that behave in ways similar to our best explanations, and to identify some of the system's logical properties.

A particularly interesting result is the fact that the *cut* rule cannot be eliminated in the system that contains a rule for abduction. This is not entirely unexpected, since abduction is typically characterized as a form of *ampliative* reasoning whose conclusions 'go beyond' the informational content of the premises. As our rule for abduction aims to capture this ampliative character, derivations that use it behave as if a new axiom were introduced midway through, prohibiting permutations of *cut*.

The rest of the paper proceeds as follows. In Section 2, we specify the particular explanatory relation we intend to represent in our object-language. Section 3 identifies seven properties that characterize our best explanations and the inferences underlying them. In Section 4, we present a sequent calculus and define a class of consequence relations within it that exhibits these properties. We begin with a base calculus for defeasible inferences that enjoys both cut-elimination and a subformula property and then combine this system with one that provides rules for explanatory argumentation. This section also features an application of the system to a canonical case study: Semmelweis' explanation of high rates of death-bed fever in Vienna's General Hospital. We then extend this calculus (Section 4.4) over a language that encodes those inferences in a *best-explains-why* operator whose elimination rule is a formalization of IBE. Finally, we prove that the extended calculus is conservative with respect to the base calculus, i.e. no sequents formulated in the old language are provable in the extension that are not already provable in the base. This means that the explanatory operator does nothing more than express or make explicit a rule of inference.

2. Explanatory Vocabulary

To start, let us specify our target. We propose a logical system whose object-language contains a connective that behaves in a manner similar to a class of natural language locutions. Paradigmatically, this class includes those expressions which are given in English as "That... best explains why...," and its nominalization "That... is the best explanation of why...," where the ellipses are filled by declarative sentences. Developing a logic attentive to the various subtleties of these natural language expressions is well beyond the scope of the current endeavor. Instead, we aim to capture at least part of the meaning of "A best explains B" as it functions in the major premise of IBE.

While many types of explanatory claims entail the truth of their explananda or their explanantia, our target notion does not. As IBE's foremost defender, Peter Lipton argues that to treat "A best explains B" as factive in the premise of IBE is "like a dessert recipe that says to start with a soufflé" (Lipton, 2004, 58). For this reason, Lipton suggests that the best explanation ought to be construed as the best *potential* explanation.[5] Following a common convention in the literature on scientific explanation, let us stipulate that "A potentially explains B" requires neither "A" nor "B" to be true.

In addition to being non-factive, the sense of *best explanation* that we aim to represent is *immediate* and *exhaustive*. We say that A is an immediate explanation of B so long as it does not explain B merely by explaining something else, C, which in turn explains B. In other words, the explanations we have in mind are not transitive. By "exhaustive," we mean that nothing needs to be added to A in order for it to explain B. The target of our account is thus expressed by the locution "That . . . is the best potential, immediate, exhaustive explanation of why . . . ". To avoid having to repeat this phrase, we will henceforth speak of *best explanations* or simply *explanations*. The connective intended to represent this expression is the binary, *best-explains-why* operator, symbolized in our formal language by '$\blacktriangleright$.'

Finally, some conceptions of *best explanation* are exclusive in the sense that their can be only one explanation that qualifies as *the best* (Bird, 2007). However, the sense of *best explanation* with which we will be working is *non-exclusive* in so far as there may be many different explanantia of the same explanandum that receive the approbation 'best'. Our decision to work with an non-exclusive conception is motivated by the fact that DIME only provides necessary conditions for best explanations.

3. Properties of Explanatory Arguments

As noted, we follow the EAI approach to explanation. In general, EAI holds that explanations are a type of inference. We refer to the inferences underlying our explanatory claims with the generic term 'explanatory argument.' Of course, we are interested in the type of inference that can be identified with our *best* explanations and are thus investigating our 'best explanatory arguments'. As we shall see, however, this evaluative feature is actualized through other properties of our consequence relation. Therefore, we will continue to identify our target as '*explanatory arguments*'. Since our aim is to have the connective '$\blacktriangleright$' exhibit the properties of our best explanations, and since we identify the latter with explanatory arguments, we must begin by specifying the properties of explanatory arguments that distinguish them from other inferences.

In order to represent explanatory arguments, we will employ the standard sequent notations—e.g. $\Gamma, A \vdash B, \Delta$ where uppercase Roman letters range over single formulas and uppercase Greek letters range over finite sets of formulas of the language of classical propositional logic. Formulas are built up from the standard connectives, $\wedge, \vee, \neg$, applied recursively to the members of a countably infinite set of atoms denoted by p, q, r. When the turnstile, $\vdash$, appears unadorned by further notation, it should be taken to denote the

[5] Lipton also argues that IBE should construe the best explanation as the "loveliest" explanation, and then vaguely describes "loveliness" as a combination of various "theoretical virtues," such as simplicity, scope, fit with background belief, etc. While DIME does appeal to a superlative—i.e. "sturdiness"—it appears to be something quite different than this. A detailed comparison of loveliness and sturdiness exceeds the scope of the current paper.

standard consequence relation for classical logic. Formulas on the left side of the turnstile are called the *antecedent*; on the right side they are called the *succedent*. Commas in the antecedent are read 'conjunctively' (i.e. as set-union) and those on the right are read 'disjunctively'. Explanatory arguments are represented by sequents with the stylized turnstile, $\overset{\blacktriangleright}{\vdash}$.

Before proceeding to the properties of explanatory arguments, let us pause to reflect upon our choice of multiple-succedent sequents to represent explanatory arguments. We read the sequent $\Gamma \overset{\blacktriangleright}{\vdash} \Delta$ as saying, in our meta-language, that Γ provides an explanatory argument for at least one member of Δ. While we do not typically talk of explanations for "at least one member of a set", this language can be retrieved from our explanatory practices. Scientists are often confronted with a range of phenomena, only some of which they can explain. In most scientific contexts, inquiry will not rest until it can be determined precisely *which* phenomena can be explained. But naturally we should not feel compelled to restrict our representation of explanatory claims to those that appear only at certain (later) stages of inquiry. Readers who are unpersuaded by these considerations will be relieved to know that our rule for IBE only operates on (and yields) single-succedent sequents.

Explanatory arguments exhibit at least seven properties. First, they are *defeasible*—an argument can cease to be explanatory when new information is added to its premises. For instance, while the claim that "The liquid's acidity explains why the blue litmus paper turned red," may constitute a good explanation, strengthening the explanans can easily produce a bad one: "The liquid's acidity and the presence of chlorine gas explains why the blue litmus paper turned red." In this case, the additional information is incompatible with the explanandum. In other cases, however, what is added to an explanans undercuts the explanatory relation itself. For instance, that the match was struck may explain why the match lit, but if it is discovered that the match was damp, then the striking no longer, by itself, explains why it lit (perhaps the match was dried before it was struck). In both sorts of cases, we say that the explanatory arguments in question are *defeated*.[6]

Definition 1 (Defeasibility)**.** $\Gamma \overset{\blacktriangleright}{\vdash} \Delta$ is *defeasible* iff $\Gamma, A \not\overset{\blacktriangleright}{\vdash} \Delta$.

Second, explanatory arguments are *minimal* in the sense that their premises provide the logically weakest explanation of their conclusion. By 'logically weakest,' we mean that no logical consequences of the premises are explanatory (of the same explanandum) unless they are logically equivalent to those premises.[7]

Definition 2 (Minimality)**.** $\Gamma \overset{\blacktriangleright}{\vdash} \Delta$ is *minimal* iff (if $\Gamma' \overset{\blacktriangleright}{\vdash} \Delta$ and $\Gamma \vdash \wedge\Gamma'$, then $\Gamma' \vdash \wedge\Gamma$).

In general, minimality prohibits the addition of irrelevant information to explanatory arguments. For example, if the liquid's acidity explains why the litmus paper turns red, then it does not follow that the liquid's acidity *and its potability* explains why the litmus paper turned red, even if the liquid is in fact potable.[8]

[6]These two forms of defeat correspond respectively to what Pollock (2015) calls *rebutting* defeaters (which provide reasons for believing the negation of the conclusion of a given inference) and *undercutting* defeaters (which challenge the support provided by the premises of a given inference). Our system does not formally distinguish between these two types of defeat.

[7]This is the standard conception of minimality in the literature on abduction (Aliseda, 2006; Gabbay & Woods, 2006; Mayer & Pirri, 1993). Other treatments of minimality appeal to preference orderings (e.g. Mayer & Pirri, 1996), however, we have already evinced our desire to avoid such extra-logical criteria.

[8]An early criticism of Hempel's covering law model was that it permitted explanations with irrelevant information, e.g. that a subject does not get pregnant is explained by the fact that he is male and takes birth control pills (Salmon & Jeffrey, 1971), or that a substance dissolves in water is explained by the fact that it is salt and has been hexed by a witch (Kyburg, 1965). By proving that the representatives of explanatory arguments in our system exhibit minimality, we thus show that DIME avoids an objection that plagued earlier EAI approaches.

Third, since contradictions do not explain, explanatory arguments must have consistent premises. Naturally, this means that the consequence relation underlying explanatory arguments must not validate the principle of *ex contradictione quodlibet* (ECQ)—i.e. $A, \neg A \vdash B$. There are numerous logical systems that renounce ECQ—most notably paraconsistent systems. However, the latter are typically designed to represent reasoning that is *tolerant* of inconsistency and that may, in some cases, even validate contradictions. By contrast, we must aim, in the opposite direction, to represent reasoning that is *intolerant* of inconsistency and contradiction. Let us call a logic that invalidates any reasoning that proceeds from inconsistent or contradictory premises *premise-consistent.* Since contradictions never explain, explanatory arguments are premise-consistent, not paraconsistent.

Definition 3 (Premise-consistency). $\Gamma \mathrel{\vdash^{\blacktriangleright}} \Delta$ is *premise-consistent* iff $\Gamma \nvdash$.

Fourth, explanatory arguments are *irreflexive.* To start, "the litmus paper turned red because it turned red" is not an explanation. More generally, no explanatory arguments should be of the form $\Gamma, A \mathrel{\vdash^{\blacktriangleright}} A$. But explanations seem to be irreflexive in an even stronger sense, for we typically treat as unacceptably 'circular' any argument whose premises contain formulas logically equivalent to one or more of those in the conclusion. Unfortunately, efforts to capture this property in a system based on classical logic are quickly frustrated. For instance, we would certainly want to reject $A, B \mathrel{\vdash^{\blacktriangleright}} B$ as circular. Since the comma that appears on the left side of the turnstile in classical sequent calculus is read conjunctively, it would seem to follow, *a fortiori*, that $A \wedge B \mathrel{\vdash^{\blacktriangleright}} B$ should also be rejected. But now consistency forces us to reject equivalent sequents such as $A \wedge (\neg A \vee B) \mathrel{\vdash^{\blacktriangleright}} B$. Pretty soon we've rejected all sequents whose antecedent follows classically from their antecedent.

It appears, then, that excluding reflexive inferences in a calculus based on classical logic requires satisfaction of a stronger constraint, namely, that explanatory arguments not be classically valid.[9] A calculus that only proves sequents that are classically invalid will be said to instantiate the property of *materiality.*

Definition 4 (Materiality). $\Gamma \mathrel{\vdash^{\blacktriangleright}} \Delta$ is *material* if only if $\Gamma \nvdash \Delta$

We thus resign ourselves to realizing materiality as a surrogate for irreflexivity. Doing so obviously curtails the derivational power of a calculus. Most notably, it means that Hempel's (1965) Deductive-Nomological (D-N) explanations, which cast explanations as valid inferences from natural laws, will not be directly represented and not simply in virtue of the expressive limitations of a propositional language. Of course, these explanations may *already* be ruled out on grounds that they violate minimality. If, for instance, $p, p \supset q \mathrel{\vdash^{\blacktriangleright}} q$ only holds when $p \mathrel{\vdash^{\blacktriangleright}} q$ does, then, since p is logically weaker than $p, p \supset q$, the first would violate minimality. So even if we were to place no restrictions on reflexivity, instances of *Modus Ponens* and, thus, of D-N explanation, might not be permitted.

Despite its failure to represent the precise D-N form, such a system does capture important features of such explanations. For instance, philosophers of science have long noted that many law-like generalizations, especially those that figure in explanations of the special sciences, admit of exceptions or only hold under certain conditions, .i.e. only *ceteris paribus* (Cartwright, 1999; Dupré, 1993; Fodor, 1974; Mitchell, 2003). By construing explanations as defeasible inferences, our account represents the non-universal character of these generalizations. In a sense, our account pushes a natural law's exceptions

[9]Redeeming our intuitions about explanatory irreflexivity might require an hyperintensional conception of reflexivity, as suggested by Jenkins (2011). For more on our view of explanation's purported hyperintensionality, see the discussion of congruentiality below.

'downward' into the explanatory relation, rather than'upward' into the generalization.

Some have argued that law-like generalizations are not even proper constituents of scientific explanations. Halonen and Hintikka (2005) claim that covering laws are at best part of the background theory against which an explanation is offered. According to Skow (2016), natural laws only give higher-level explanations, e.g. explaining why one event caused another. Since our calculus reserves a role for background information and permits conditionals to serve in explanations, it could redeem either of these insights. At base, these authors advocate a conceptual distinction between a statement that explains and the law-like generalization in virtue of which it explains. Our account perspicuously captures this distinction.

If, however, one finds that the loss of valid inferences from the pool of potential explanatory arguments too high a cost to bear, one may opt instead to relax the materiality constraint. Doing so in our calculus is quite easy and we explain how to accomplish it in the paper's conclusion. For purposes of exposition, we have chosen to present the system in its material guise, leaving it up to the reader to decide whether to retain this restriction.

As will become clear, our characterization of explanatory arguments requires comparison of defeasible inferences that are premise-consistent and material. In order to facilitate this comparison, we introduce the following definition.

Definition 5 (Non-Triviality)**.** An inference $\Gamma \mid\!\sim \Delta$ is *non-trivial* iff it is defeasible, premise-consistent, and material.

Fifth, explanations are *stable*, which philosophers of science have analyzed in different ways (Hempel, 1965; Lange, 2009; Mitchell, 2003; Skyrms, 1980; Woodward, 2003). In its most general form, X is said to be stable if X remains unchanged as other conditions C change. For instance, suppose that a patient's rash is explained by a particular bacterial infection, though an alternative potential explanation is that the rash is caused by an allergic reaction. The explanation is stable insofar as she would have a rash regardless of whether she had had an allergic reaction. Typically, the fundamental bearers of stability are taken to be laws or generalizations, but for us they are explanatory arguments themselves.

One of the hallmarks of DIME is the distinctive brand of stability that it attributes to our best explanations—a comparative property among defeasible inferences that we call *sturdiness*.[10] An inference is sturdy just in case it succeeds when all other non-trivial inferences that share its conclusion fail, where by 'failure' we mean *defeat* and by 'success', the lack of defeat. To test a candidate inference for sturdiness, we take its set of competitors, i.e. the logically weakest or *minimal* premise-sets that non-trivially imply its conclusion, and we suppose that each member of each premise-set is false. Recall that premise-consistent, defeasible inferences are defeated when information that yields inconsistency is added to their premises. Since competitors are premise-consistent, adding the negation of their premises defeats them. Thus, the test of sturdiness allows us to see whether the candidate remains undefeated under conditions that defeat its competitors (i.e. when the latter's premises are false).

Suppose, for example, that there are only three candidate explanatory arguments for the patient's rash, namely that she has a bacterial infection, that she has an allergic reaction, and that she has dry skin. If the inference from "the patient has a bacterial infection" to "the patient has a rash" remains undefeated under assumptions that "the patient does not have an allergic reaction" and "the patient does not have dry skin," then it is sturdy. In this sense, we like to think of sturdiness as a *test* that we perform on inferences.

[10]For a more thorough treatment of sturdiness, see Khalifa et al. (2018a).

As noted above, we aim to codify the logic of a non-exclusive conception of *best explanation*. We observe that there is nothing in the slogan for *sturdiness* that prohibits more than one set of formulas qualifying as sturdy with respect to a given explanandum. In other words, there is nothing to rule out *ties* for the status of best explanation. Such a scenario occurs when two or more sequents remains undefeated when all of the (premises of the) competitors are false.

Sixth, explanations are *detachable*. If the bacterial infection is the best explanation of the rash, then we may conclude that the patient has a bacterial infection. By detaching the explanans, we may make further predictions, design interventions, and construct new models. This is the animating idea behind IBE. Unlike non-evaluative forms of abduction for which the conclusion is thought to hold only provisionally, the conclusion of IBE is supposed to have as much warrant as the premises.

Lastly, the explanatory relations operative in science appear to be congruential in the sense that they persist under substitution of logical equivalents.

Definition 6 (Congruentiality)**.** If $\bigwedge\Gamma \dashv\vdash \bigwedge\Gamma'$ and $\bigvee\Delta \dashv\vdash \bigvee\Delta'$, then $\Sigma \mid \Gamma \mathrel{|\!\!\overset{\blacktriangleright}{\underset{\Theta}{}}} \Delta$ iff $\Sigma \mid \Gamma' \mathrel{|\!\!\overset{\blacktriangleright}{\underset{\Theta}{}}} \Delta'$.

Schnieder (2011) claims that 'because' is not a congruential expression and that it is in fact hyperintensional—i.e. necessarily equivalent clauses are not substitutable *salva veritate*. However, as Schnieder acknowledges, his claims concern the use of 'because' to express so-called metaphysical explanations or *grounding* relations. To our knowledge, no one has argued that *scientific* explanation is hyperintensional. Moreover, intuitions supporting the purported hyperintensionality of explanation appear stronger at the sub-sentential level—i.e. when co-intensional names or predicates are intersubstituted—than at the sentential level. Compare: "the liquid's acidity explains why the litmus paper turned red" with "the liquid's acidity explains why the litmus paper turned the color of Susan's favorite dress" and "the liquid's acidity explains why it was not the case that the litmus paper did not turn red." Nevertheless, we remain agnostic on the general question of whether scientific explanations are fundamentally extensional, intensional, or hyperintensional. Since we have chosen to work with the operators of classical logic, congruentiality seems to be an appropriate desideratum. It is entirely possible, however, that this property is an artifact of the propositional context. We make no assumptions about whether it holds for explanatory arguments formulated in richer languages.

4. The Logic of Explanatory Arguments and IBE

Below, we develop sequent calculi for explanatory arguments, IBE, and the *best-explains-why* operator. We introduce a logic, $\mathsf{LEA}_{\mathbb{P}}$, defined over the standard propositional language, $\mathcal{L}$, whose sequents represent rules of inference in the sense that anyone who is entitled to the formulas in the antecedent is entitled to those in the succedent. The *rules* of our calculi, on the other hand, tell us what rules of inference are legitimate given the axioms—i.e. they are meta-rules of inference.

$\mathsf{LEA}_{\mathbb{P}}$ is composed of two parts: $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ and $\mathsf{LE}^{\blacktriangleright}$. The first part, $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$, is a variant of Smullyan's 'symmetric' sequent calculus for classical logic.The consequence relations operative in $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ are relativized both to a *defeater set* that represents the defeaters of the inference and to a set of *proper axioms*, $\mathbb{P}$, that represents the non-logical inferences taken for granted in a particular scientific domain. $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ exhibits many desirable properties for sequent calculi, including cut-elimination, decidability, and variants of the uniformity and subformula properties. The second part of the base system, $\mathsf{LE}^{\blacktriangleright}$ introduces an additional class of sequents and rules that govern their interaction with those of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$, including a turnstile-level rule for IBE. Later, we describe a system $\mathsf{LEA}^{+}_{\mathbb{P}}$ defined over an extension of $\mathcal{L}$ that includes an object-language connective, $\blacktriangleright$, intended to make commitment to explanatory arguments explicit. As will become clear, throughout this work we assume the soundness, completeness, and decidability of SC.

The need for a base logic that includes two types of sequents follows from the fact that if the *best-explains-why* operator encodes explanatory arguments, and if IBE is itself an explanatory argument, then for any valid instance of IBE, there ought to be a theorem of the form *[(A best explains why B)* $\wedge$ *B] best explains why A*. But, insofar as sentences of this form are even intelligible, it is far from clear whether they *claim* what an application of IBE *shows*. Thus, IBE is not an explanatory argument. Rather, IBE belongs to one class of inferences and explanatory arguments belong to another.

In recognition of this point, $\mathsf{LEA}_{\mathbb{P}}$ contains two consequence relations, or, more precisely, two *classes* of consequence relations, for as we shall see, the system contains $\mathsf{card}(\mathcal{P}(\mathcal{L}))$-many consequence relations. That class whose rules are given by $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ is intended to represent the type of inferences to which both IBE as well as *candidate* explanatory arguments belong. The class of consequence relations introduced by $\mathsf{LE}^{\blacktriangleright}$ and denoted by $\mathrel{|\!\!\overset{\blacktriangleright}{-}}$ captures the behavior of explanatory arguments themselves.

In what follows, we assume a propositional language, $\mathcal{L}$, for classical logic, that consists of a countably infinite set of atomic sentences $\mathsf{Atoms} = \{p, q, r, \ldots, p_n, q_n, r_n\}$ and the standard connectives $\wedge, \vee, \neg$. As before, formulas are defined in the usual manner and are ranged over by A, B, C, etc. Literals are formulas that are either atoms or their negations and are denoted by l, l_n. The set of literals in $\mathcal{L}$ is denoted by Lit. Sets of formulas are ranged over by $\Gamma, \Delta, \Theta, \Psi, \Sigma, \Lambda, \Xi$; sets of literals by X, Y, and sets of sets of literals by $\mathbf{S}, \mathbf{T}$. All of the latter sets are assumed to be finite. Finally, we will let $\mathbb{P}$ denote finite sets of sequents that meet certain specifications.

The sequents in our calculi depart from the standard form in two respects: just below our turnstile we add a set of formulas, Θ, called a *defeater set*, and to the far left-hand side we add another, Σ, called a *background set*. As with antecedents in standard sequents with sets, use the comma to denote set union in the background set, i.e. $\Sigma \cup \{A\} = \Sigma, A$.

$$\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\sim}} \Delta$$

Defeater sets 'describe' situations in which one is not permitted to draw the inference represented by the sequent. Roughly put, a sequent is defeated whenever the defeater set is a logical consequence of the set composed of its background and antecedent. Thus, as the name suggests, defeater sets are sets of inference-defeaters.

We interpret background sets as consisting of information that is available to a reasoner when she draws an inference, but which does not serve as a premise and from which the conclusion is not said to follow. Having such a device in our formalism enables us to capture an important aspect of defeasibility, namely, that the introduction of new information may jeopardize prior inferential commitments even when that information does not serve as fodder for new inferences. Moreover, it allows us to highlight the role that background information plays in the evaluation of explanatory claims.[11]

[11] It is customary for logic-based approaches to abduction to begin with the definition of an *abduction problem*. In its simplest form, an abduction problem consists of the following task: Given a *background theory* Γ, a fact B, and a specified consequence relation, $\mathrel{|\!\sim}$, find a formula, A, such that $\Gamma \mathrel{|\!\not\sim} B$ but $\Gamma, A \mathrel{|\!\sim} B$. When using the notation of defeasible sequents to construct such problems, it is important not to confuse *background sets* with *background theories*. In our representation of explanatory arguments, background sets are intended to represent information that is not part of an exhaustive explanation. This information need not be veridical, e.g. when potential explanantia rely upon idealized assumptions. So, while background information need not be true, but it must not be 'up for grabs.' In this sense, placing theoretical commitments in the background set is a way of representing the status of such commitments in Kuhnian "normal science." On the other hand, in periods of what Kuhn calls "revolutionary science," such commitments may become part of the reasoning

Because the provability of any sequent in $\mathsf{LEA}_{\mathbb{P}}$ depends, in part, upon the contents of its defeater set, there is not one or two but $\mathsf{card}(\mathcal{P}(\mathcal{L}))$-many consequence relations represented by the calculi. Some are classical, i.e. $\Theta = \emptyset$, but many are non-monotonic, i.e. $\Theta \neq \emptyset$. By specifying the contents of defeater sets, the rules of our calculi are able to exploit this panoply so as to home in on the class of consequence relations that bears the properties of explanatory arguments.

Definition 7 (Defeater sets, Background sets, Defeasible Sequents)**.** Defeater sets are finite, possibly empty sets of formulas that defeat an inference (see below). Background sets are finite, possibly empty sets of formulas that represent the background assumptions of the inference. A defeasible sequent is a standard sequent in which a *background set* (Σ) occurs on the left-hand side of the antecedent and a *defeater set* (Θ) occurs just below the turnstile:[12] $\Sigma \mid \Gamma \mathrel{|}\!\!\!\underset{\Theta}{\sim} \Delta$. When no background sets have been specified (i.e. $\Sigma = \emptyset$) we write: $\cdot \mid \Gamma \vdash_{\Theta} \Delta$. When the antecedent is empty we write: $\Sigma \mid \cdot \vdash_{\Theta} \Delta$.

Definition 8 (Defeat)**.** A defeasible sequent is said to be *defeated* just in case the defeater set is a (nonempty) LK-consequence of the antecedent-*cum*-background set, i.e.

$$\Sigma \mid \Gamma \vdash_{\Theta} \Delta \quad \text{is defeated } \textit{iff} \quad \Sigma, \Gamma \vdash \Theta' \text{ for some } \emptyset \subset \Theta' \subseteq \Theta.$$

Example 1. $\Sigma \mid \Gamma, p \wedge q \vdash_{\{p\}} \Delta$ is defeated.

Example 2. $\Sigma \mid \Gamma, p \vdash_{\{p \vee q\}} \Delta$ is defeated.

Example 3. $\cdot \mid p \vdash_{\{p \wedge q\}} \Delta$ is undefeated.

Example 4. $\cdot \mid p \vee q \vdash_{\{p\}} \Delta$ is undefeated.

The definition of *defeat* is designed to redeem certain intuitions about defeasible inferences. For instance, if we know that A defeats an inference, then we ought to reject that inference if we are entitled to $A \wedge B$ (Example 1). Conversely, if $A \vee B$ defeats an inference, then the inference is defeated if we are entitled to either of the disjuncts (Example 2). If we know that $A \wedge B$ defeats an inference, but are ignorant as to whether either conjunct by itself defeats it, then we should not abandon the inference if we are merely entitled to one of the conjuncts (Example 3). In this sense, the use of conjunctions in the defeater set permits a reasoner to flag instances when the combination of certain pieces of information undermines an inference even when neither would on its own. Lastly, it seems that if A defeats an inference, then entitlement to $A \vee B$ need not force us to abandon that inference, since entitlement to B would preserve the propriety of the inference (Example 4).

From our formulation of *defeat*, we can identify sequents whose corresponding inferences are defeasible, premise-consistent, and material. For instance, it is obvious that any sequent whose defeater set is nonempty will be defeasible.

Lemma 1. *If* $\Theta \neq \emptyset$*, then* $\Sigma \mid \Gamma \vdash_{\Theta} \Delta$ *is defeasible.*

process proper, in which case they can be moved to the antecedent. In Khalifa et al. (2018a) we show how, when combined with the sturdiness test, the manipulation of background sets represents the way scientists' theoretical and practical interests constrain the range of potential explanations.

[12] In general, the respective roles played by *repositories*, *control sets* and *soundness* in Piazza and Pulcini's $\mathsf{LK}^{\mathcal{S}}$ are played by *background sets*, *defeater sets* and *undefeatedness* in $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$. However, we have made extensive changes to the system in light of explanation-relevant considerations and our fundamentally different notion of *defeat* (resp. *soundness*).

Proof. If $\Theta \neq \emptyset$ then $\exists A \in \Theta$ such that $\Sigma, \Gamma, A \vdash \Theta$ and thus from Definition 8 it follows that $\Sigma \mid \Gamma, A \not\vdash_{\Theta} \Delta$, satisfying Definition 1. ■

One of the reasons for formulating defeat in terms of nonempty consequences is that inconsistent antecedent-*cum*-background sets entail even the empty set, and thus sequents with such inconsistencies would not be provable (there is no proof of a defeated sequent in our calculi). But sequents with inconsistent antecedents are often required to prove theorems of classical logic, and we would like to include as many of these in our system as possible. By defining defeat as we have, inconsistent antecedents/backgrounds are permitted when the defeater set is empty. Thus, inconsistency only leads to defeat when the defeater set is nonempty.

Lemma 2. *If $\Theta \neq \emptyset$ and $\cdot \mid \Gamma \vdash_{\Theta} \Delta$ is undefeated, then $\cdot \mid \Gamma \vdash_{\Theta} \Delta$ is premise-consistent.*

Proof. Suppose for *reductio* that $\Gamma \vdash$. It follows that $\Gamma \vdash \Theta$ for any $\Theta \subseteq \mathcal{L}$. But then the sequent is defeated, contradicting the hypothesis. So the sequent satisfies Definition 3. ■

Lastly, material (i.e. classically non-valid) inferences are easily represented by defeasible sequents whose defeater sets contain their succedents.

Lemma 3. *If $\emptyset \neq \Delta \subseteq \Theta$ and $\Sigma \mid \Gamma \vdash_{\Theta} \Delta$ is undefeated, then $\Gamma \nvdash \Delta$.*

Proof. By hypothesis, it follows via Definition 8 that $\Sigma, \Gamma \nvdash \Delta$. ■

4.1 $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$

The rules for $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ are given in Figure 1. The new or relocated formula in the conclusion of a rule is called *principal*, its subformulas in the premises are called *active*, and the remaining elements of sequents are called *side formulas*. The rules may be read (top-down) as entitling a reasoner to endorse the conclusion-inference, given her entitlement to the premises. As Definition 9 implies, the rules allow a defeated sequent to be obtained from undefeated premises, but the result is a *paraproof*, not a proof.

Definition 9 (Derivation, Proof, Paraproof in $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$)**.** A *derivation* is a rooted, finitely branching tree $\mathcal{D}$ whose nodes are sequents of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ and which is recursively built up from axioms by means of the rules of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$. If each sequent in $\mathcal{D}$ is undefeated, then $\mathcal{D}$ is said to be a *proof* of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$, otherwise $\mathcal{D}$ is called a *paraproof* of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$.

$\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ is based on Smullyan's (1968) SC for classical logic. The calculus employs 'symmetric' rules for logical connectives to ensure that an end-sequent's subformulas will be preserved on the same side of the turnstile throughout the derivation.[13] In order to achieve completeness, the system pushes the behavior LK's negation rules into the logical axioms, which explains why the system has three forms for such axioms. Our axiomatized version of SC extends this treatment to the set of proper axioms by closing the latter under a kind of contraposition.

$\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ differs from Smullyan's SC in several respects. To start, SC has no structural rules, not even *cut*, while our $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ enjoys rules for both weakening and cut. Since our system retains the core rules of Smullyan's, our proof of cut-elimination may readily be

[13]Smullyan presents his calculus in terms of the distinction between α and β formulas. The former are formulas of the forms $\neg\neg A, A \wedge B, \neg(A \vee B)$ or $\neg(A \supset B)$, while β formulas are $A \vee B, \neg(A \wedge B)$ or $A \supset B$. Left rules for α formulas and right rules for β formulas are called *A-rules* while the converse pair are called *B-rules*. The A-rules are unary, while the B-rules are binary.

Logical Axioms

$$\frac{}{\cdot \mid p \vdash_{\emptyset} p}\ \textit{log. ax.} \qquad \frac{}{\cdot \mid p, \neg p \vdash_{\emptyset}}\ \textit{log. ax.} \qquad \frac{}{\cdot \mid \cdot \vdash_{\emptyset} p, \neg p}\ \textit{log. ax.}$$

Proper Axioms

$$\frac{}{\Sigma \mid X \vdash_{\Theta} Y}\ \textit{prop. ax.}^{\dagger}$$

Cut Rule

$$\frac{\Sigma \mid \Gamma \vdash_{\Theta} A, \Delta \qquad \Sigma' \mid \Gamma', A \vdash_{\Psi} \Delta'}{\Sigma', \Sigma \mid \Gamma', \Gamma \vdash_{\Theta \cup \Psi} \Delta, \Delta'}\ \textit{cut}$$

Structural Rules

$$\frac{\Sigma \mid \Gamma \vdash_{\Theta} \Delta}{\Sigma \mid \Gamma, A \vdash_{\Theta} \Delta}\ \mathsf{LW} \qquad \frac{\Sigma \mid \Gamma \vdash_{\Theta} \Delta}{\Sigma \mid \Gamma \vdash_{\Theta} A, \Delta}\ \mathsf{RW}$$

$$\frac{\Sigma \mid \Gamma \vdash_{\Theta} \Delta}{\Sigma \mid \Gamma \vdash_{\Theta \cup \Psi} \Delta}\ \mathsf{DE} \qquad \frac{\Sigma \mid \Gamma \vdash_{\Theta} \Delta}{\Sigma, A \mid \Gamma \vdash_{\Theta} \Delta}\ \mathsf{BE}$$

Logical Rules

$$\frac{\Sigma \mid \Gamma, A, B \vdash_{\Theta} \Delta}{\Sigma \mid \Gamma, A \wedge B \vdash_{\Theta} \Delta}\ \wedge\vdash \qquad \frac{\Sigma \mid \Gamma \vdash_{\Theta} A, \Delta \qquad \Sigma' \mid \Gamma' \vdash_{\Psi} B, \Delta'}{\Sigma', \Sigma \mid \Gamma', \Gamma \vdash_{\Theta \cup \Psi} A \wedge B, \Delta, \Delta'}\ \vdash\wedge$$

$$\frac{\Sigma \mid \Gamma, A \vdash_{\Theta} \Delta \qquad \Sigma' \mid \Gamma', B \vdash_{\Psi} \Delta'}{\Sigma', \Sigma \mid \Gamma', \Gamma, A \vee B \vdash_{\Theta \cup \Psi} \Delta, \Delta'}\ \vee\vdash \qquad \frac{\Sigma \mid \Gamma \vdash_{\Theta} A, B, \Delta}{\Sigma \mid \Gamma \vdash_{\Theta} A \vee B, \Delta}\ \vdash\vee$$

$$\frac{\Sigma \mid \Gamma, A \vdash_{\Theta} \Delta}{\Sigma \mid \Gamma, \neg\neg A \vdash_{\Theta} \Delta}\ \neg\neg\vdash \qquad \frac{\Sigma \mid \Gamma \vdash_{\Theta} A, \Delta}{\Sigma \mid \Gamma \vdash_{\Theta} \neg\neg A, \Delta}\ \vdash\neg\neg$$

$$\frac{\Sigma \mid \Gamma, \neg A \vdash_{\Theta} \Delta \qquad \Sigma' \mid \Gamma', \neg B \vdash_{\Psi} \Delta'}{\Sigma', \Sigma \mid \Gamma', \Gamma, \neg(A \wedge B) \vdash_{\Theta \cup \Psi} \Delta, \Delta'}\ \neg\wedge\vdash \qquad \frac{\Sigma \mid \Gamma \vdash_{\Theta} \neg A, \neg B, \Delta}{\Sigma \mid \Gamma \vdash_{\Theta} \neg(A \wedge B), \Delta}\ \vdash\neg\wedge$$

$$\frac{\Sigma \mid \Gamma, \neg A, \neg B \vdash_{\Theta} \Delta}{\Sigma \mid \Gamma, \neg(A \vee B) \vdash_{\Theta} \Delta}\ \neg\vee\vdash \qquad \frac{\Sigma \mid \Gamma \vdash_{\Theta} \neg A, \Delta \qquad \Sigma' \mid \Gamma' \vdash_{\Psi} \neg B, \Delta'}{\Sigma', \Sigma \mid \Gamma', \Gamma \vdash_{\Theta \cup \Psi} \neg(A \vee B), \Delta, \Delta'}\ \vdash\neg\vee$$

† Provided that $\langle \Sigma, X, \Theta, Y \rangle \in \mathbb{P}$.

Figure 1.: Rules for $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$

applied to SC. To our knowledge, we are the first to provide a cut-elimination theorem for Smullyan's calculus.

Another point of difference is that our $\mathcal{L}$ does not contain an expression for the material conditional. Instead, the rules for disjunction can be applied when no more than one of the active formulas is negated. The rules for negated formulas thus require the main operator in *each* active formula to be negation. Naturally, $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ includes proper axioms as well as special structural rules for defeasible sequents (DE, BE). Finally, where SC permits logical axioms with arbitrary side formulas, $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ does not. Instead, $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ has rules for left and right weakening in order to permit weakening in derivations containing proper axioms, which obviously cannot be formulated with arbitrary side formulas.[14]

There are other features of the calculus worthy of note. For instance, with the exception of DE, unary (i.e. one-premise) rules have no effect on defeater sets, whereas binary (i.e. two-premise) rules yield defeater sets in the conclusion that are the union of those in the premises. This fact guarantees that no information about potential defeaters is lost along a derivation. The presence of DE is grounded in the recognition that reasoners ought to be able to add new, extra-logical information about defeaters to their arguments as it becomes available. The DE rule (which stands for *Defeater Expansion*) allows one to do so as long as it does not defeat the sequent in question.

Similarly, all of the rules either transfer or combine background sets from premises to conclusions, except for BE. The former (whose name abbreviates *Background Expansion*) permits the addition of arbitrary formulas to a sequent's background set. At one level, this rule simply captures the way new contextual information is added in the course of scientific reasoning. But at a deeper level, it attempts to represent the way reasoners might *probe* the defeasibility of an inference by discharging it in different contexts. In this sense, BE codifies certain patterns of *experimental* reasoning.

$\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ is defined by a finite set of proper axioms, denoted by $\mathbb{P}$, whose members are represented by 4-tuples of the following form: $\langle \Sigma, X, \Theta, Y \rangle$, where Σ is the background set, X is the antecedent, Θ is the defeater set, and Y the succedent. Proper axioms represent the non-logical starting point of scientific reasoning. Since such inferences are typically the products of concrete empirical inquiries, we insist that they be introduced with non-empty background and defeater sets of (possibly complex) formulas that reflect the context of their use. It is imagined that the set of proper axioms will change depending upon the domain and broader context in which reasoning occurs.

Not just any sequent composed of literals should be allowed to qualify as the starting point of explanatory reasoning. For reasons already discussed, proper axioms must also be premise-consistent, material, and defeas*ible*, yet not defeat*ed*. Naturally, the empty sequent should not be provable from the members of $\mathbb{P}$, for this would render any cut-free non-logical axiomatic extension of classical propositional logic inconsistent.[15] The background set of any proper axiom ought to be *substantive* in the sense that it provides unique information regarding the context of inference. In order to permit logically equivalent proper axioms in a system that lacks tradition negation rules, we must ensure that literals can be removed from one side of the turnstile and replaced by their complement on the other—affecting a kind of closure under contraposition. Finally, in order to permute *cut* to the top of derivations, and hence to ensure cut-elimination, $\mathbb{P}$ should be closed under the *cut* rule. We impose the following constraints on proper axioms to secure these conditions.

Definition 10 (Constraints on Members of $\mathbb{P}$)**.** Let $\mathsf{Cn}(\Gamma)$ be the closure of Γ under the rules of classical logic and let $\mathbb{C}$ be the closure of $\{\Sigma \mid X \mid\!\!\sim_{\Theta} Y : \langle \Sigma, X, \Theta, Y \rangle \in \mathbb{P}\}$

---

[14]Our decision to adopt Smullyan's SC rather than Gentzen's LK as the basis for the defeasible calculus was made on largely technical grounds. The rules for negation in LK permit formulas to traverse the turnstile. Since defeat is defined with respect to formulas on the left-hand side of the turnstile, the behavior of such rules makes it far more difficult to track defeat throughout a derivation and thus complicates normalization. A similar calculus that *is* based on Gentzen's LK is presented in Millson (2018), where it is used to provide a proof theory for Inferential Erotetic Logic

[15]It is well-known that the addition of non-tautological axioms to the system of classical logic will lead to inconsistency if those axioms are taken to be closed under universal substitution (US). Even if closure under US is abandoned for proper axioms, their addition to a sequent system such as LK, threatens the cut-elimination theorem. Fortunately, Piazza and Pulcini (2016) have shown how to generate non-logical axiomatic extensions of classical propositional logic that admit cut elimination. While such extensions are obviously not complete—they are *post-complete*—axioms can be formulated so as to preserve consistency. The trick to doing so is to ensure that the empty sequent does not belong to the set of proper axioms (see Theorem 3.7 in Piazza and Pulcini (2016)). Our proper axioms have been formulated in conformity with this constraint.

under *cut*. We then require that, for any $\Sigma \subset \mathcal{L}$ and $X, Y \subset \mathsf{Lit}$ and $\Theta \subset \mathcal{P}(\mathsf{Lit})$,

I. If $\langle \Sigma, X, \Theta, Y \rangle \in \mathbb{P}$, then
   (i) $\Sigma \mid X \mathrel{\vert\!\!\sim_{\Theta}} Y$ is not defeated;
   (ii) if $X(\text{resp.}\, Y) = \emptyset$, then $Y(\text{resp.}\, X) \neq \emptyset$.
   (iii) $\Sigma \neq \emptyset \neq \Theta$; and.
   (iv) $\mathsf{Cn}(\Sigma) \cap \mathsf{Cn}(Y) = \mathsf{Cn}(\emptyset) = \mathsf{Cn}(X) \cap \mathsf{Cn}(Y)$.

II. $\langle \Sigma, X \cup \{l\}, \Theta, Y \rangle \in \mathbb{P}$ iff $\langle \Sigma, X, \Theta, Y \cup \{l^c\} \rangle \in \mathbb{P}$

III. $\mathbb{P} = \big\{ \langle \Sigma, X, \Theta, Y \rangle : \Sigma \mid X \mathrel{\vert\!\!\sim_{\Theta}} Y \in \mathbb{C} \big\}$.

The first item in constraint (I) is straightforward. The second excludes the empty sequent from $\mathbb{P}$. The third guarantees defeasibility via Lemma 1. The fourth achieves two goals. First, it prohibits logical entailment between antecedents and succedents, thereby securing materiality and premise-consistency. Second, when combined with (iii), it guarantees that the background set of a proper axiom provides at least one unique piece of contextual information. Constraint (II) closes the set under contraposition and (III) closes it under *cut*. We note that these constraints permit the background and defeater sets of proper axioms to contain complex formulas. Leaving this possibility open provides flexibility in the representation of initial inferences.

$\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ also comes equipped with a rule for introducing standard logical axioms. As noted above, in order to permit various logical inferences to be provable in the calculus, we must allow sequents with inconsistent antecedents to be derived from the logical axioms. Since $A, \neg A \vdash \Theta$ for any $\Theta \subset \mathcal{L}$, we require logical axioms to have empty defeater sets. This stipulation means that $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ preserves the theorems of classical logic.

Lemma 4. $\Sigma \mid \Gamma \mathrel{\vert\!\!\sim_{\emptyset}} \Delta$ *iff* $\Gamma \vdash \Delta$.

Proof. Since proper axioms must have nonempty defeater sets, we know that a provable sequent with an empty defeater set must either be a logical axiom or be derivable from logical axioms alone. Our result thus follows from the fact that the rules for $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ are just those of SC (*modulo* background and defeater sets) which is adequate for classical logic. ■

Lemma 5. *Any sequent that is provable in* $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ *has a cut-free proof.*

Proof. See Appendix A. ■

As we will see, $\mathsf{LE}^{\blacktriangleright}$ contains a rule whose application assumes that there is a terminating algorithm which determines for each sequent of $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ whether it is or is not provable. We must therefore prove the decidability of $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$.

Proposition 6. $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ is decidable.

Proof. We begin by restricting our attention to cut-free derivations in $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ in which no sequent appears twice in a branch, following the usual reduction to *concise* proofs. When confronted with an end-sequent $\Sigma \mid \Gamma \mathrel{\vert\!\!\sim_{\Theta}} \Delta$, we must first determine whether it is defeated. If it is, then there is no proof. If it is undefeated, then we proceed in the manner of the standard *proof search algorithm* for sequent calculi, i.e. we consider all the possible inference rules that could have $\Sigma \mid \Gamma \mathrel{\vert\!\!\sim_{\Theta}} \Delta$ as a conclusion and construct a number of trees, one for each distinct possibility, writing down the premises of such rules above $\Sigma \mid \Gamma \mathrel{\vert\!\!\sim_{\Theta}} \Delta$. Matters are slightly complicated by the presence of BE and DE in $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$. However, since the background and defeater sets of any sequent are finite, we know that there is an upper bound to the number of derivations for any end-sequent containing a nonempty background or defeater set. We then repeat this process for the premises of each successive rule, at each point checking to see whether the premises are defeated. If a premise is defeated in a tree, then, following Definition 9, it is not a possible proof, and thus we turn to the remaining trees. We proceed in this fashion until we have trees with no defeated sequents whose leaves are either logical axioms or sequents whose antecedents and succedents are sets of literals. If all of these initial sequents are either logical axioms or members of $\mathbb{P}$ we have found our desired proof. If not, then the sequent is not provable in $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$. ■

We can see that $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ is congruential, as expected.

Lemma 7. *If* $\bigwedge \Gamma \dashv\vdash \bigwedge \Gamma'$ *and* $\bigvee \Delta \dashv\vdash \bigvee \Delta$, *then* $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{-}} \Delta$ *is provable in* $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ *iff* $\Sigma \mid \Gamma' \mathrel{|\!\!\underset{\Theta}{-}} \Delta'$ *is.*

Proof. We proceed by cases. For our first case, suppose $\Theta = \emptyset$. The lemma naturally holds for LK *modulo* background and defeater set notation. Thus, from Lemma 4, we obtain the result for sequents with empty defeater sets. Now consider the (second) case in which the defeater sets are nonempty, i.e. $\Theta \neq \emptyset$. If $\bigwedge \Gamma \dashv\vdash \bigwedge \Gamma'$ or $\bigvee \Delta \dashv\vdash \bigvee \Delta$, then, via Lemma 4, it follows that $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\emptyset}{-}} \Gamma'$; $\Sigma \mid \Gamma' \mathrel{|\!\!\underset{\emptyset}{-}} \Gamma$; $\Sigma \mid \Delta \mathrel{|\!\!\underset{\emptyset}{-}} \Delta'$; and $\Sigma \mid \Delta' \mathrel{|\!\!\underset{\emptyset}{-}} \Delta$. We obtain the biconditional by applying *cut* to (the relevant) one of these sequents, e.g. $\Sigma \mid \Gamma' \mathrel{|\!\!\underset{\emptyset}{-}} \Gamma$ and $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{-}} \Delta$ yields $\Sigma \mid \Gamma' \mathrel{|\!\!\underset{\Theta}{-}} \Delta$ via iterated applications of *cut* to the formulas in Γ. ■

Since explanatory arguments will be drawn from sequents provable in $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$, and since explanation supervenes upon relevance relations of one sort or another, the calculus ought to satisfy some recognizable relevance criteria. We now show that the system exhibits what we call *basic* $\mathbb{P}$*-relevance*, an augmented version of the *basic relevance criterion* that Avron (2016) attributes to the 'semi-relevant' logic **RM**. Our *basic* $\mathbb{P}$*-relevance* criterion combines Avron's basic relevance criterion with a relevance-tracking property specific to proper axiom sets. Basic $\mathbb{P}$-relevance is defined in terms of the rather straightforward notion of *valence.*

Definition 11. The *valence*, both positive and negative, of an occurrence of an atom p in a formula A is defined recursively as follows.

- p occurs positively in p
- If p occurs positively (negatively) in A, then it occurs negatively (positively) in $\neg A$
- If p occurs positively (negatively) in A or B, then it occurs positively (negatively) in $A \wedge B$ and $A \vee B$.

Definition 12. Let $\mathsf{Val}(\Sigma) = \{p : p \text{ occurs positively in some } A \in \Sigma\} \cup \{\neg p : p \text{ occurs negatively in some } A \in \Sigma\}$.

Definition 13. We say l is $\mathbb{P}$*-relevant* to l' iff there is a $\langle \Sigma, X, \Theta, Y \rangle \in \mathbb{P}$ for which $l \in X$ and $l' \in Y$ or if $\mathsf{Atoms}(l) \cap \mathsf{Atoms}(l') \neq \emptyset$. We write $\Gamma \mid_{\mathbb{P}} \Gamma'$ iff for all $l \in \mathsf{Val}(\Gamma)$ and all $l' \in \mathsf{Val}(\Gamma')$, l is not $\mathbb{P}$-relevant to l'.

Fact 1. $\mid_{\mathbb{P}}$ is symmetric and irreflexive but not necessarily transitive.

Fact 2. $\Gamma \mid_{\emptyset} \Gamma'$ iff $\mathsf{Atoms}(\Gamma) \cap \mathsf{Atoms}(\Gamma') = \emptyset$.

Sequent-calculi that are adequate for classical logic satisfy the following uniformity property: If $\Gamma \cup \Delta \mid_{\emptyset} \Gamma'$; $\Gamma' \nvdash$; and $\Gamma, \Gamma' \vdash \Delta$, then $\Gamma \vdash \Delta$. Avron (2016) defined the following basic relevance criterion, which is, for instance, satisfied by the logic **RM**: If $\Gamma \cup \Delta \mid_{\emptyset} \Gamma'$ and $\Gamma, \Gamma' \vdash \Delta$ then $\Gamma \vdash \Delta$. (Here we've adjusted it to our notation and multi-conclusion consequence relations.) In Appendix B, we show that $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$ satisfies basic $\mathbb{P}$-relevance and thus also uniformity *modulo* $\mathbb{P}$.

Proposition 8 (Basic $\mathbb{P}$-Relevance of $\mathsf{SC}^{\mathbf{\Theta}}_{\mathbb{P}}$)**.** If $\Theta \neq \emptyset$, $\Sigma \mid \Gamma, \Gamma' \mid\!\frac{}{\Theta}\, \Delta$ is provable, and $\Gamma \cup \Delta \mid_{\mathbb{P}} \Gamma'$, then $\Sigma \mid \Gamma \mid\!\frac{}{\Theta}\, \Delta$ is also provable.

Finally, in Appendix B, we show that $\mathsf{SC}^{\mathbf{\Theta}}_{\mathbb{P}}$ exhibits a version of the standard subformula property for sequent calculi. More precisely, we show that the following holds.

Proposition 9 (Subformula Property)**.** Where $\mathcal{D}$ is a cut-free derivation of $\Sigma \mid \Gamma \mid\!\frac{}{\Theta}\, \Delta$,

(1) if $\neg A$ occurs in Γ' $[\Delta']$ for some sequent $\Sigma' \mid \Gamma' \mid\!\frac{}{\Theta'}\, \Delta'$ in $\mathcal{D}$, then $\neg A$ occurs positively in Γ $[\Delta]$ or A occurs negatively in Γ $[\Delta]$.
(2) where A is a non-negated formula, if A occurs in Γ' $[\Delta']$ for some sequent $\Sigma' \mid \Gamma' \mid\!\frac{}{\Theta'}\, \Delta'$ in $\mathcal{D}$, then A occurs in Γ $[\Delta]$.

4.2 LE$^{\blacktriangleright}$

With the rules for $\mathsf{SC}^{\mathbf{\Theta}}_{\mathbb{P}}$ now in place, we can turn to the other part of our base system, $\mathsf{LE}^{\blacktriangleright}$, where we introduce a class of consequence relations, denoted by $\mid\!\frac{\blacktriangleright}{\Theta}$, that represents explanatory arguments. We will call sequents constructed with this turnstile $\blacktriangleright$-sequents. The crucial property of provable $\blacktriangleright$-sequents is that of *sturdiness*, which, as mentioned above, is DIME's proposal for how to understand the general property of *stability* associated with explanations.

According to DIME, explananda are the sturdy consequences of explanantia. Recall our slogan for sturdiness: inferences are sturdy just in case *they succeed where all their competitors fail.* We have already begun to unpack this claim; a sturdy inference is one that remains undefeated under any of the conditions that falsify its competitors—i.e. the premises of non-trivial inferences that share its conclusion. To express this slogan in our meta-language, we need first to generate a precise definition of *competitor.* In what follows, let us call the antecedent of an inference for which we are interested in generating a competitor set a *candidate for sturdiness* or simply a *candidate.*

Definition 14 (Antecedent Set, $\mathbf{S}_{\langle \Sigma, \Delta \rangle}$)**.** For any $\emptyset \subset X \subset \mathsf{Lit}$ and $\emptyset \subset \Sigma, \Theta, \Delta \subset \mathcal{L}$, let $\mathbf{S}_{\langle \Sigma, \Delta \rangle} =_{df} \left\{ X : \Sigma \mid X \mid\!\frac{}{\Theta \cup \Delta}\, \Delta \text{ is provable in } \mathsf{SC}^{\mathbf{\Theta}}_{\mathbb{P}} \right\}$.

Despite its complex appearance, $\mathbf{S}_{\langle \Sigma, \Delta \rangle}$ simply contains all those sets of literals that serve as antecedents in defeasible, premise-consistent, and material sequents provable in $\mathsf{SC}^{\mathbf{\Theta}}_{\mathbb{P}}$ that have Δ as their succedent and Σ as their background set. Note that it is the fact that the succedent, i.e Δ is included in the defeater sets of these sequents that guarantees their materiality. Naturally, the computability of antecedent sets presupposes the decidability of $\mathsf{SC}^{\mathbf{\Theta}}_{\mathbb{P}}$, which we obtain from Proposition 6.

While restricting the antecedent set to sets of literals ensures that the competitor set is finite, doing so naturally curtails the expressive range of competitors. Note, however, that it does not compromise logical strength, as every logically complex formula is the consequence of a set of literals.

Nonetheless, in light of the minimality of explanations, some constraints on the logical strength of competitors will be needed. For among the members of any antecedent set are sets of literals built up from iterated applications of LW. This opens up the possibility that the set union of the antecedent set will contain complementary literals. Consider $\mathbf{S}_{\langle \Sigma, \Delta \rangle} = \{\{p\}, \{p, q\}, \{p, \neg q\}\}$. In this case, $\bigcup \mathbf{S}_{\langle \Sigma, \Delta \rangle}$ is inconsistent. Of course, such inconsistencies ought only to arise from the addition of superfluous formulas. As we shall soon see, in rendering the competitor set from the antecedent set, we need to apply set union to the latter, negate its elements, and test for inconsistency with the candidate. Therefore, we must cull sets with superfluous premises that might give rise to inconsistencies in advance of comparison with the candidate. To do so, we reduce the antecedent set to its logically weakest, or, *minimal* members.

Definition 15 (Literal-Set Minimization, $\mathbf{S}^{\downarrow}$)**.** For any $X, Y \subset \mathsf{Lit}$ and $\mathbf{S} \subset \mathcal{P}(\mathsf{Lit})$, let $\mathbf{S}^{\downarrow} =_{df} \mathbf{S} \backslash \{X : Y \in \mathbf{S}, Y \subset X\}$

Example 5. $\{\{p\}, \{p, q\}, \{p, \neg q\}\}^{\downarrow} = \{\{p\}\}$

Example 6. $\{\{p, q, r\}, \{p, \neg q\}, \{p, r\}\}^{\downarrow} = \{\{p, \neg q\}, \{p, r\}\}$

This operation yields a set whose members are *minimal* in the sense that none is logically weaker than any other. Since we are dealing exclusively with sets of literals, this property is realized in terms of set inclusion, i.e. no set is a proper subset of any other.

Of course, the minimized antecedent set for any background-succedent pair is not yet *a candidate's* competitor set for the simple reason that it makes no reference to a candidate. In order to arrive at a candidate's proper competitor set, we must cull elements that competitors share with the candidate.

Naturally, a candidate for sturdiness should not compete against itself. Nor should it compete against a set of its literal consequences unless that set is logically weaker than it and belongs to the minimized antecedent set. Thus, to generate a set of proper competitors for some candidate Γ while ensuring that successful candidates are minimal, we must only remove Γ's literal consequences from those competitors in $\mathbf{S}^{\downarrow}_{\langle\Sigma,\Delta\rangle}$ that are not logically weaker than Γ. Definition 18 illustrates this careful removal procedure. But first, we introduce some useful auxiliary definitions.

Definition 16 (Logically weaker than, $\Gamma >^{\vdash} \Delta$)**.** $\Gamma >^{\vdash} \Delta$ iff $\Gamma \vdash \bigwedge \Delta$ but $\Delta \nvdash \bigwedge \Gamma$. When $\Gamma >^{\vdash} \Delta$, we say that Δ is logically weaker than Γ.

Definition 17 (Literal consequence-set, $\mathsf{Cn}_{\mathsf{Lit}}(\Gamma)$)**.** $\mathsf{Cn}_{\mathsf{Lit}}(\Gamma) =_{df} \{l \in \mathsf{Lit} : \Gamma \vdash l\}$

Definition 18 ($\mathbf{S} \setminus\!\!\setminus \Gamma$)**.** For any $X, Y \subset \mathsf{Lit}, \Gamma \subset \mathcal{L}$ and $\mathbf{S} \subset \mathcal{P}(\mathsf{Lit})$, let

$$\mathbf{S} \setminus\!\!\setminus \Gamma =_{df} \bigcup\{X : X \in \mathbf{S}, \Gamma >^{\vdash} X\} \cup \Big(\big(\bigcup\{Y : Y \in \mathbf{S}, \Gamma \not>^{\vdash} Y\}\big) \setminus \mathsf{Cn}_{\mathsf{Lit}}(\Gamma)\Big)$$

Example 7. $\{\{\neg p\}, \{q\}, \{p, r\}\} \setminus\!\!\setminus \{p, q\} = \{q\} \cup \big(\{\neg p, p, r\} \backslash \mathsf{Cn}_{\mathsf{Lit}}(\{p, q\})\}\big) = \{q, \neg p, r\}$

Example 8. $\{\{p\}, \{q\}, \{\neg p, \neg r\}\} \setminus\!\!\setminus \{p \wedge q\} = \{p, q\} \cup \big(\{\neg p, \neg r\} \backslash \mathsf{Cn}_{\mathsf{Lit}}(\{p \wedge q\})\}\big)$
$= \{p, q, \neg p, \neg r\}$

Example 9. $\{\{p, q, r_1\}, \{p, q, \neg r_2\}, \{\neg r_3\}\} \setminus\!\!\setminus \{p \vee q, r_1\}$
$= \emptyset \cup \big(\{p, q, r_1, \neg r_2, \neg r_3\} \backslash \mathsf{Cn}_{\mathsf{Lit}}(\{p \vee q, r_1\})\}\big) = \{p, q, \neg r_2, \neg r_3\}$

We at last have the means to precisely define what counts as a set of competitors.

Definition 19 (Competitor Set, $\Xi_{\langle\Sigma,\Gamma,\Delta\rangle}$)**.** For any sequent, $\Sigma \mid \Gamma \mathrel{|\!\!\!\sim_{\Theta}} \Delta$, the competitor set of its antecedent, Γ, is $\Xi_{\langle\Sigma,\Gamma,\Delta\rangle} =_{df} \mathbf{S}^{\downarrow}_{\langle\Sigma,\Delta\rangle} \setminus\!\!\setminus \Gamma$.

Notice that it is consistent with this definition that a candidate's competitor set include not only inferences that we would recognize as *potential* explanations—e.g. if the match was struck, then it lit—but also decidedly non-explanatory, evidential inferences—e.g. if a flame was observed, then the match lit. Moreover, all members of the antecedent set and thus the members of a candidate's competitor set must belong to defeasible sequents that are provable under the same set of background conditions as that of the candidate, i.e. Σ. This establishes a common set of assumptions against which comparisons can be made.

Since the explanandum is a non-trivial consequence of each competitor, and since non-trivial inferences are defeated when *any* of their premises (members of their antecedents) are false (negated), we represent all those situations in which a candidate's competitors *fail* as the situation in which every member of the competitor set is false, i.e. $\neg\Xi_{\langle\Sigma,\Gamma,\Delta\rangle} = \neg(\mathbf{S}^{\downarrow}_{\langle\Sigma,\Delta\rangle} \setminus\!\!\setminus \Gamma)$.

To see whether a candidate for sturdiness succeeds amidst the failure of its competitors, we simply add this set to the background set of the candidate sequent. If the resultant sequent is undefeated, it succeeds where its competitors fail and thus passes the test of sturdiness. If it is defeated, it fails and is thus not the best explanation.

Since the formalization of the sturdiness test requires that competitor sets be added to the backgrounds of defeasible sequents, we must be certain that competitor sets are finite, even though Lit is infinite.

Proposition 10. For some $n \in \mathbb{N}$, $\mathsf{card}(\Xi_{\langle\Sigma,\Gamma,\Delta\rangle}) = n$.

Proof. Note that if $\Sigma \mid X \,\big|\!\!\underline{}_{\Theta\cup\Delta}\, \Delta$ is provable then by basic $\mathbb{P}$-relevance $\Sigma \mid X' \,\big|\!\!\underline{}_{\Theta\cup\Delta}\, \Delta$ is also provable where $X' \subseteq X$ and every $l \in X'$ is $\mathbb{P}$-relevant to some $l' \in \mathsf{Val}(\Delta)$. From this it follows that for every $\Xi \in \mathbf{S}^{\downarrow}_{\langle\Sigma,\Delta\rangle}$, $\Xi \subseteq \mathsf{RelLit}(\Delta)$ where $\mathsf{RelLit}(\Delta) = \{l : l$ is $\mathbb{P}$-relevant to some $l' \in \mathsf{Val}(\Delta)\}$.

Note also that $\mathsf{RelLit}(\Delta) = \{l : \mathsf{Atoms}(l) \cap \mathsf{Atoms}(\Delta) \neq \emptyset\} \cup \{l : \exists\langle\Sigma', X', \Theta', Y'\rangle \in \mathbb{P}$ for which $l \in X'$ and $Y' \cap \mathsf{Val}(\Delta) \neq \emptyset\}$. Since Δ is finite, $\{l : \mathsf{Atoms}(l) \cap \mathsf{Atoms}(\Delta) \neq \emptyset\}$ is finite and since $\mathbb{P}$ is finite, $\{l : \exists\langle\Sigma', X', \Theta', Y'\rangle \in \mathbb{P}$ for which $l \in X'$ and $Y' \cap \mathsf{Val}(\Delta) \neq \emptyset\}$ is finite as well. So, also $\mathsf{RelLit}(\Delta)$ is finite. In view of this and $\Xi \subseteq \mathsf{RelLit}(\Delta)$ it follows that $\mathbf{S}^{\downarrow}_{\langle\Sigma,\Delta\rangle}$ is finite. ■

We will now consider the rules for $\mathsf{LE}^{\blacktriangleright}$, presented in Figure 2. The rule we call STR attempts to capture the test of sturdiness we have just described. This rule serves as a kind of 'gatekeeper' for $\blacktriangleright$-sequents, only allowing through those sequents of $\mathsf{SC}^{\Theta}_{\mathbb{P}}$ that are non-trivial and sturdy. Non-triviality is achieved by the stipulation that the rule's first premise must be undefeated and must have a (nonempty) defeater set that includes its succedent. The sturdiness test is performed by the second premise, which is identical to the first except for the addition of the negated competitor set to the sequent's background. Again, the sequent must be undefeated to apply the rule. The test-like character of STR lies in the fact that the negated competitor set does not carry down into the conclusion. This behavior mandates a binary formulation of the rule. Were the rule to contain only the second premise, it might be possible to derive $\blacktriangleright$-sequents that had no provable counterparts in $\mathsf{SC}^{\Theta}_{\mathbb{P}}$. For instance, suppose that $\neg p \mid q \,\big|\!\!\underline{}_{\Theta}\, \Delta$ is provable but $r \mid q \,\not\big|\!\!\underline{}_{\Theta}\, \Delta$. Via BE, we have $\neg p, r \mid q \,\big|\!\!\underline{}_{\Theta}\, \Delta$. Suppose further that $\neg\Xi_{\langle\{r\},\{q\},\Delta\rangle} = \neg p$. A unary STR would permit the derivation of $r \mid q \,\big|\!\!\underline{}^{\blacktriangleright}_{\Theta}\, \Delta$ despite the fact that $r \mid q \,\not\big|\!\!\underline{}_{\Theta}\, \Delta$.

As the name suggests, the ABD rule is intended to capture the detachability of the premises of explanatory arguments via *abductive* inference. Roughly put, ABD says that if Γ and A best explains Δ, and Γ' provides evidence for Δ, then together, Γ and Γ' license the inference to A. Since A only forms part of the explanatory argument for Δ, we may read ABD as licensing the detachment of a *partial* explanation of Δ—though still the 'best' partial explanation. The formulation of the rule thus makes detachability a manifest property of $\blacktriangleright$-sequents.

There are, however, some peculiarities to the rule that deserve discussion. First, the explanandum (Δ) disappears from the conclusion, leaving only the partial explanans. This feature accords with our desire to present IBE in the strongest form possible. If IBE only licensed inferences to best explanations *or* their explananda, it's legitimacy would hardily have roused debate—though its utility might have. Unfortunately, the absence of the explananda in the succedent of ABD's conclusion means that information is lost

Mixed Rules

$$\frac{\Sigma \mid \Gamma \vdash_{\Theta \cup \Delta} \Delta \qquad \neg\Xi_{\langle \Sigma, \Gamma, \Delta \rangle},\ \Sigma \mid \Gamma \vdash_{\Theta \cup \Delta} \Delta}{\Sigma \mid \Gamma \vdash^{\blacktriangleright}_{\Theta \cup \Delta} \Delta}\ \mathsf{STR}^{\S}$$

$$\frac{\Sigma \mid \Gamma, A \vdash^{\blacktriangleright}_{\Theta} \Delta \qquad \Sigma' \mid \Gamma' \vdash_{\Psi \cup \Delta} \Delta}{\Sigma', \Sigma \mid \Gamma', \Gamma \vdash_{\Theta \cup \Psi \cup \Delta} A}\ \mathsf{ABD}^{\P}$$

Logical Rules

$$\frac{\Sigma \mid \Gamma, A \vdash^{\blacktriangleright}_{\Theta} \Delta}{\Sigma \mid \Gamma, \neg\neg A \vdash^{\blacktriangleright}_{\Theta} \Delta}\ \neg\neg\vdash^{\blacktriangleright} \qquad \frac{\Sigma \mid \Gamma \vdash^{\blacktriangleright}_{\Theta} A, \Delta}{\Sigma \mid \Gamma \vdash^{\blacktriangleright}_{\Theta} \neg\neg A, \Delta}\ \vdash^{\blacktriangleright}\neg\neg$$

$$\frac{\Sigma \mid \Gamma, A, B \vdash^{\blacktriangleright}_{\Theta} \Delta}{\Sigma \mid \Gamma, A \wedge B \vdash^{\blacktriangleright}_{\Theta} \Delta}\ \wedge\vdash^{\blacktriangleright} \qquad \frac{\Sigma \mid \Gamma \vdash^{\blacktriangleright}_{\Theta} A, B, \Delta}{\Sigma \mid \Gamma \vdash^{\blacktriangleright}_{\Theta} A \vee B, \Delta}\ \vdash^{\blacktriangleright}\vee$$

$$\frac{\Sigma \mid \Gamma, \neg A, \neg B \vdash^{\blacktriangleright}_{\mathbf{S}} \Delta}{\Sigma \mid \Gamma, \neg(A \vee B) \vdash^{\blacktriangleright}_{\mathbf{S}} \Delta}\ \neg\vee\vdash^{\blacktriangleright} \qquad \frac{\Sigma \mid \Gamma \vdash^{\blacktriangleright}_{\Theta} \neg A, \neg B, \Delta}{\Sigma \mid \Gamma \vdash^{\blacktriangleright}_{\Theta} \neg(A \wedge B), \Delta}\ \vdash^{\blacktriangleright}\vee$$

§ Provided that (i) the premise is undefeated and (ii) $\Sigma, \Gamma, \Theta, \Delta \neq \emptyset$.

¶ Provided that $\Gamma, A \nvdash \Gamma'$.

Figure 2.: Rules for $\mathsf{LE}^{\blacktriangleright}$

in any derivation that contains an application of the rule. The effect, like that of proofs that employ *cut*, is the failure of analyticity—i.e. derivations may contain formulas that are not subformulas of those in the end-sequent.

Another issue of note is that the second, evidential premise of ABD is restricted to sequents whose defeater set contains the succedent. This restriction ensures that the premise is empirically or materially informed. Similarly, the proviso (¶) on the rule prevents the evidential premise from containing logical consequences of the explanans. In this way, we accommodate the idea that abductive inferences are only licensed when there is evidence for the explanandum that is independent of the explanans. It follows that the second premise of ABD will contain a sequent whose antecedent (or a set of literals logically equivalent to it) may have 'competed' with the explanans for sturdiness. This is as it should be, since, as we have noted, the competitor set includes not only potential explanations, but also non-explanatory, evidential inferences. Of course, nothing in the rule precludes the antecedents of the second, evidential premise in ABD from being empty and therefore representing a material or empirical fact.

The STR rule encodes counterfactual reasoning: would the candidate remain good if it's competitors were false? The ABD rule, on the other hand, considers whether a competitor's *actuality* provides evidence of the explanandum's obtaining. Competitors play distinct rules in each rule. Consider: the striking of the match would cause the match to light regardless of whether a flame was observed. But that the flame *was* observed provides evidence that the flame lit, whose best explanation is that it was struck.

The logical rules of $\mathsf{LE}^{\blacktriangleright}$ are similar to the corresponding rules in $\mathsf{SC}_{\mathbb{P}}^{\mathbf{\Theta}}$, *modulo* $\blacktriangleright$-sequents. Only the unary rules for connectives are permitted in $\mathsf{LE}^{\blacktriangleright}$. (These represent what Smullyan calls *A-rules*, see note 13 above.) The reason for this restriction is that the binary rules involve the 'fusion' of side-formulas in a manner that does not seem permissible in explanatory reasoning. For instance, in general, explanations of two distinct phenomena cannot be combined to yield a satisfactory explanation of the conjunction of the two. Consider: that it is the first Monday of the month best explains why the sirens are sounding (its a regular test of the warning system) and that it is autumn best explains why there is smoke in the air (residents are burning leaves), but the fact that it is the first Monday of the month and that it is autumn does not best explain why the sirens are sounding and their is smoke in the air (there's been an explosion!). As this example illustrates, combining the best explanations of two facts need not yield the best explanation of their conjunction, especially when their conjunction is far less likely than their independent occurrence. Moreover, combination of arbitrary antecedents often produces explanatory arguments that are no longer premise-consistent or minimal—as is readily observed in the case of $\vee\vdash$. Insofar as the rules for the omitted connectives play a role in explanatory arguments, these propositions need to be built up through applications of the rules of $\mathsf{SC}_{\mathbb{P}}^{\mathbf{\Theta}}$, not through manipulation of $\blacktriangleright$-sequents.

Perhaps most striking in $\mathsf{LE}^{\blacktriangleright}$ is the absence of classical structural rules. Neither weakening nor cut is permitted. If the antecedents of $\blacktriangleright$-sequents could be arbitrarily weakened, then the minimality condition, which STR secures, would be compromised. Right weakening, on the other hand, does not directly conflict with the properties of explanatory arguments; rather, we deny it because the weakening of explananda by arbitrary disjuncts appears to us as both unnatural and unreasonable. Finally, the absence of *cut* follows from the fact that our target vocabulary is that which expresses the notion of *immediate* explanation.

4.3 $\mathsf{LEA}_{\mathbb{P}}$

Now that we have in place its constituent systems, we must define derivations in $\mathsf{LEA}_{\mathbb{P}}$. This definition will depart from that of $\mathsf{SC}_{\mathbb{P}}^{\mathbf{\Theta}}$ because the conclusion of ABD cannot serve as a candidate explanatory argument. The reason for this is that, as noted above, instances of IBE are not themselves explanatory arguments. We thus restrict the application of STR to those sequents that are not derived via STR itself.

Nor, however, should the conclusion of ABD be allowed to serve as a premise in any further application of that rule. Permitting these sorts of derivations would lead to abductive chains that assume transitive explanations. Thus, if we are to represent *immediate* explanations, symmetry demands that we prohibit transitive abduction. Moreover, chained abductions appear to be significantly weaker, epistemically speaking, then those for which there is non-explanatory evidence of an explanandum's obtaining; the longer the chain of abductions the more *speculative* their nature. As our goal is to represent scientific reasoning, such chainings should be prohibited.

In order to prevent instances of IBE from serving as explanatory arguments and to prohibit chained abduction, we insist that proofs in $\mathsf{LEA}_{\mathbb{P}}$ (and its extension, $\mathsf{LEA}_{\mathbb{P}}^{+}$) only be permitted one application of each of the following: $\mathsf{STR}, \mathsf{ABD}$ (and $\vdash\blacktriangleright$, $\blacktriangleright\vdash$ for $\mathsf{LEA}_{\mathbb{P}}^{+}$). We can now define derivations in $\mathsf{LEA}_{\mathbb{P}}$. This definition will also hold for derivations in the extended system $\mathsf{LEA}_{\mathbb{P}}^{+}$ *modulo* the latter's additional rules.

Definition 20 (Derivation, Proof, Paraproof in $\mathsf{LEA}_{\mathbb{P}}(\mathsf{LEA}_{\mathbb{P}}^{+})$)**.** A *derivation* is a rooted, finitely branching tree $\mathcal{D}$ whose nodes are sequents of $\mathsf{SC}_{\mathbb{P}}^{\mathbf{\Theta}}$-*cum*-$\mathsf{LE}^{\blacktriangleright}$, i.e. $\mathsf{LEA}_{\mathbb{P}}$, which is recursively built up from axioms by means of the rules of both, but in which *no branch contains more than one occurrence of a rule in the set* $\{\mathsf{STR}, \mathsf{ABD}\}$. *Proof* and *paraproof* are the same as those given in Definition 9. For the extension $\mathsf{LEA}^{+}_{\mathbb{P}}$, derivations are defined in terms of the rules given in Figure 3 and branches must have no more than one application of the following rules: $\{\mathsf{STR}, \mathsf{ABD}, \vdash\blacktriangleright, \blacktriangleright\vdash\}$. Again, *proof* and *paraproof* are the same as those given in Definition 9.

4.3.1 A Worked Example

Let's see how derivations are constructed in $\mathsf{LEA}_{\mathbb{P}}$. To illustrate, we'll use the well-worn example of Ignaz Semmelweis' attempt to explain the high occurrence of death by childbed fever exhibited by patients in the First Maternity Division of Vienna General Hospital between 1844 and 1848. First elaborated by Hempel (1966, 3–8) and later discussed at length by Lipton (2004, 74–90), this example has come to represent a canonical instance of both scientific explanation and IBE. According to Hempel, Semmelweis sought to explain why the rate of death by childbed fever in the First Division was substantially greater than that in the Second. He considered various explanations before concluding that the cause was the presence of "cadaveric matter" on the hands and instruments of surgeons who often performed deliveries immediately after conducting autopsies. One rejected hypothesis was that "atmospheric-cosmic-telluric changes" in the districts from which patients hailed were responsible for an epidemic of childbed fever. However, as Semmelweis observed, such an explanation could not account for the relatively low number of cases of the disease that occurred in the Second Division. Moreover, there was in fact no epidemic of childbed fever in Vienna at the time. Another plausible explanation was the fact that women in the First Division delivered on their backs—in what is called the lithotomy position—to accommodate the (male!) surgeons, while those in the Second delivered on their sides—lateral position—under the care of midwives. However, after instituting lateral-position delivery in the First Division, Semmelweis observed no change in the rate of infection. Other hypotheses were tested—such as that the presence of a priest in the First Division's delivery ward had terrifying and debilitating effects on patients—until Semmelweis ordered surgeons to disinfect their hands prior to delivering infants. The result was a substantial reduction in the cases of childbed fever.

Lipton (2004) used this example as a case of IBE and emphasized the way in which contrasts in the explanadum, i.e. between death rates in the First and Second Divisions, figured prominently. However, for simplicity, we won't represent these contrasts. Doing so would not be difficult, but would require more variables and sequents. Instead, we shall focus on the explanations of high rates of childbed fever in the First Division. We further simplify the case by considering only the "cadaveric matter" hypothesis (p_2), the lithotomy-position hypothesis (q_1), and the "priest" hypothesis (q_2), which, incidentally, were the last three tested by Semmelweis. Since the "epidemic" hypothesis had by now been ruled out, we'll take this information to form the background of the inferences ($\neg r_1$). We'll also assume that the chief evidence for the explanandum was the observation of symptoms associated with the disease in patients who expired (r_2). The following, then, are the interpretations of the atoms involved in the derivation and its set of proper axioms.

p_1 = Surgeons perform deliveries.
p_2 = Cadaveric matter is present on the hands and instruments of surgeons.
q_1 = Women deliver on their backs.
q_2 = A priest is present in the delivery ward.
r_1 = There is an epidemic of childbed fever.
r_2 = A significant number of patients exhibit symptoms of childbed fever.
r_3 = A significant number of patients die of childbed fever.

$$\mathbb{P} = \left\{ \begin{array}{l} \langle \{\neg r_1\}, \{p_1, p_2\}, \{\neg p_2\}, \{r_3\} \rangle \text{ [cadaveric-matter hypothesis]}, \\ \langle \{\neg r_1\}, \{p_1, q_1\}, \{\neg p_2, \neg q_1\}, \{r_3\} \rangle \text{ [lithotomy-position hypothesis]}, \\ \langle \{\neg r_1\}, \{q_2\}, \{\neg p_2, \neg q_2\}, \{r_3\} \rangle \text{ [priest hypothesis]}, \\ \langle \{\neg r_1\}, \{r_2\}, \{\neg p_2, \neg r_2\}, \{r_3\} \rangle \text{ [evidential hypothesis]} \end{array} \right\}$$

These axioms are not intended to capture Semmelweis' idiosyncratic conception of each hypothesis, but rather to represent inferential relations and background, e.g. causal, information that we can now, retrospectively, see as relevant to his experimental interventions. For instance, we have included the fact that surgeons perform deliveries (p_1) in the lithotomy hypothesis because it is was the surgeons, not the mothers, who preferred that position and had the power to dictate its use. Whether Semmelweis appreciated this point is orthogonal to the task of characterizing the relevant candidates for best explanation of childbed fever. More strikingly, notice that the absence of cadaveric matter defeats both the inference from lithotomy position and the inference from the presence of a priest. The fact that sterilizing surgeons' hands and instruments undermines the propriety of either inference would only have been recognized by Semmelweis after he had hit upon the best explanation. In other words, axioms are formulated so as to provide a retrospective reconstruction of explanatory reasoning, not to facilitate a prospective *discovery* of best explanations.

Example 10. Proof of $\neg r_1 \mid r_2 \vdash_{\{\neg p_2, \neg r_2, r_3\}} p_1 \wedge p_2$.

$$
\begin{prooftree}
\AxiomC{}
\RightLabel{\scriptsize $prop.ax.$}
\UnaryInfC{$\neg r_1 \mid p_1, p_2 \vdash_{\{\neg p_2\}} r_3$}
\RightLabel{\scriptsize DE}
\UnaryInfC{$\neg r_1 \mid p_1, p_2 \vdash_{\{\neg p_2, r_3\}} r_3$}
\AxiomC{}
\RightLabel{\scriptsize $prop.ax.$}
\UnaryInfC{$\neg r_1 \mid p_1, p_2 \vdash_{\{\neg p_2\}} r_3$}
\RightLabel{\scriptsize DE}
\UnaryInfC{$\neg r_1 \mid p_1, p_2 \vdash_{\{\neg p_2, r_3\}} r_3$}
\RightLabel{\scriptsize BE}
\UnaryInfC{$\neg q_1, \neg q_2, \neg r_1, \neg r_2 \mid p_1, p_2 \vdash_{\{\neg p_2, r_3\}} r_3$}
\RightLabel{\scriptsize STR}
\BinaryInfC{$\neg r_1 \mid p_1, p_2 \overset{\blacktriangleright}{\vdash}_{\{\neg p_2, r_3\}} r_3$}
\RightLabel{\scriptsize $\wedge\vdash^{\blacktriangleright}$}
\UnaryInfC{$\neg r_1 \mid p_1 \wedge p_2 \overset{\blacktriangleright}{\vdash}_{\{\neg p_2, r_3\}} r_3$}
\AxiomC{}
\RightLabel{\scriptsize $prop.ax.$}
\UnaryInfC{$\neg r_1 \mid r_2 \vdash_{\{\neg p_2, \neg r_2\}} r_3$}
\RightLabel{\scriptsize DE}
\UnaryInfC{$\neg r_1 \mid r_2 \vdash_{\{\neg p_2, \neg r_2, r_3\}} r_3$}
\RightLabel{\scriptsize ABD}
\BinaryInfC{$\neg r_1 \mid r_2 \vdash_{\{\neg p_2, \neg r_2, r_3\}} p_1 \wedge p_2$}
\end{prooftree}
$$

The derivation's center branch contains an application of BE that adds denials of the lithotomy-position hypothesis ($\neg q_1$), the priest hypothesis ($\neg q_2$), and the evidential hypothesis ($\neg r_2$) to the sequent's background set. The first two denials represent Semmelweis' effort to institute lateral-delivery and to remove the priest from the maternity ward. The fact that the cadaveric-matter hypothesis "survives" the addition of this information is then represented by the conclusion of STR in which the $\blacktriangleright$-turnstile is introduced. Recall that $\blacktriangleright$-sequents stand for best-explanatory arguments. Thus, the STR-step in the proof represents the identification of the cadaveric-matter hypothesis as the best explanation.

The axioms associated with the rival hypotheses would not survive were the surgeons to disinfect before performing deliveries ($\neg p_2$), since $\neg p_2$ belongs to both of their defeater sets. We observe that the proof's end-sequent inherits the defeater set of the evidential inference, which includes the denial that symptoms of childbed fever were observed. Consequently, the instance of IBE that concludes the proof is sensitive to information that might undermine our confidence that the explanandum obtains. Since our commitment to an IBE should only extend as far as we are warranted in believing the explanandum, this strikes us as both intuitive and desirable.

Before proceeding to a discussion of our main results, let us note that a slight modification to the contents of $\mathbb{P}$ would produce a case in which more than one $\blacktriangleright$-sequent with r_3 for its succedent would be provable. For instance, if the disinfection of surgeon's hands, i.e. $\neg p_2$, were removed from the defeater set of the lithotomy-position hypothesis, then it too would pass the sturdiness test. This could be established through a separate derivation or by appending a branch to the one above if a conjunctive explanation were sought. Such a scenario would represent the situation in which more than one hypothesis qualified as the best explanation. This fits with our *non-exclusive* interpretation of that concept. In our discussion below, we provide a relevant example in which there is an explanatory 'tie' between incompatible hypotheses.

4.3.2 Main Results

We are now in a position to show that provable $\blacktriangleright$-sequents exhibit many of the characteristics we identify with our best explanations. We begin with those of defeasibility, premise-consistency, and materiality.

Lemma 11. *If $\Sigma \mid \Gamma \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$, then $\emptyset \subset \Delta \subseteq \Theta$.*

Proof. Follows from the formulation of STR. ∎

Corollary 12. *If the sequent $\Sigma \mid \Gamma \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$, then it is defeasible.*

Proof. From Lemmas 1 and 11. ∎

Corollary 13. *If the sequent $\Sigma \mid \Gamma \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$, then it is premise-consistent.*

Proof. From Lemmas 2 and 11. ∎

Corollary 14. *If the sequent $\Sigma \mid \Gamma \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$, then it is material.*

Proof. From Lemmas 11 and 3. ∎

We can also show that $\blacktriangleright$-sequents are minimal in the sense given by Definition 2. To prove minimality, we proceed by proving some useful lemmas, the first of which shows that the logical rules of $\mathsf{LE}^{\blacktriangleright}$ are admissible in $\mathsf{LEA}_{\mathbb{P}}$.

Lemma 15. *The logical rules for $\mathsf{LE}^{\blacktriangleright}$ are admissible in $\mathsf{LEA}_{\mathbb{P}}$, i.e. every sequent that can be derived using the rule is already derivable without it.*

Proof. We prove that $\wedge\vdash^{\blacktriangleright}$ and $\vdash^{\blacktriangleright}\vee$ are admissible by induction on proof-height, leaving the proof for the remaining rules as an exercise for the reader.

($\wedge\vdash^{\blacktriangleright}$) Suppose that $\mathcal{D}$ is a proof of $\Sigma \mid \Gamma \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ whose last rule is $\wedge\vdash^{\blacktriangleright}$. Since the shortest proof of this sort applies $\wedge\vdash^{\blacktriangleright}$ to the conclusion of STR, we take as our base case $\mathcal{D}$ such that $\Sigma \mid \Gamma' \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ obtains via STR. It follows that $\Sigma \mid \Gamma' \vert\!\underset{\Theta}{-} \Delta$ must occur in the premises of STR and hence that it is provable. By supposition, there are some formulas $A, B \in \Gamma'$ such that $A \wedge B \in \Gamma$. We thus apply $\wedge\vdash$ to this sequent to obtain $\Sigma \mid \Gamma \vert\!\underset{\Theta}{-} \Delta$. Via $\mathsf{card}(\Xi_{\langle\Sigma,\Gamma,\Delta\rangle})$-many applications of BE we obtain the premise of STR. Since $\forall X(\{A, B\} >^{\vdash} X \Leftrightarrow \{A \wedge B\} >^{\vdash} X)$, it follows from Definition 19 that $\Xi_{\langle\Sigma,\Gamma',\Delta\rangle} = \Xi_{\langle\Sigma,\Gamma,\Delta\rangle}$ and thus from our supposition, $\Sigma \mid \Gamma \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ is obtained via STR. Our result is then achieved by induction on the height of $\mathcal{D}$.

($\vdash^{\blacktriangleright}\vee$) By Definition 19, $\Xi_{\langle\Sigma,\Gamma,\{A,B\}\cup\Delta\rangle} = \Xi_{\langle\Sigma,\Gamma,\{A\vee B\}\cup\Delta\rangle}$. The result follows straightforwardly. ∎

We may now see that every provable $\blacktriangleright$-sequent has a provable counterpart in $\mathsf{SC}^{\Theta}_{\mathbb{P}}$.

Lemma 16. *If $\Sigma \mid \Gamma \vert\!\overset{\blacktriangleright}{\underset{\Theta}{-}} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$, then $\Sigma \mid \Gamma \vert\!\underset{\Theta}{-} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$.*

Proof. Provable ►-sequents must contain STR in their derivation and may also contain applications of the logical rules. If $\Sigma \mid \Gamma \overset{\blacktriangleright}{\underset{\Theta}{\mid\!\!\sim}} \Delta$ follows directly from STR, then, given the formulation of that rule, $\Sigma \mid \Gamma \mid\!\!\sim_{\Theta} \Delta$ must occur in its premises and hence must be provable. Our result holds regardless of any further application of logical rules, since, according to Lemma 15, such rules are admissible. ■

The following Fact 3 shows that explanantia which contain disjunctions always imply one of their disjuncts. In this sense, disjunctions in explanantia never introduce ambiguity or incompleteness. Indeed, as Fact 4 states, best explanantia are always equivalent to a conjunction of their literal consequences. Irreducibly disjunctive explanations are rightly regarded as *incomplete*, and our system excludes them. These results are thus reassuring to the extent that they cohere with our target concept of complete or *exhaustive* explanations.

Definition 21 (Minimal DNF)**.** A formula, $\bigvee_{i=1}^{n} A_i$, in disjunctive normal form (DNF) is minimal iff there is no $1 \leq j \leq n$ such that $\bigvee\{A_i : i \in \{1, \ldots, j-1, j+1, \ldots n\}\} \vdash \bigvee_{i=1}^{n} A_i$.

Fact 3. We show that if $\Sigma \mid \bigvee_{i=1}^{n} A_i \overset{\blacktriangleright}{\underset{\Theta}{\mid\!\!\sim}} \Delta$ is provable and $\bigvee_{i=1}^{n} A_i$ is a minimal DNF, then $n = 1$.

Proof. Suppose $\Sigma \mid \bigvee_{i=1}^{n} A_i \mid\!\!\sim_{\Theta} \Delta$ is provable where $\bigvee_{i=1}^{n} A_i$ is a minimal DNF and $A_i = \bigwedge \Gamma_i$. Assume for a contradiction that $n \geq 2$. By applying *cut* to $A_i \mid\!\!\sim_{\emptyset} \bigvee_{i=1}^{n} A_i$ and $\Sigma \mid \bigvee_{i=1}^{n} A_i \mid\!\!\sim_{\Theta} \Delta$, $\Sigma \mid A_i \mid\!\!\sim_{\Theta} \Delta$ is derivable. We first show that there is at least one $i \in \{1, \ldots, n\}$ for which $\Sigma \mid A_i \mid\!\!\sim_{\Theta} \Delta$ is not defeated. Assume the opposite. Then $\Sigma, A_i \vdash \Theta$ for each $i \in \{1, \ldots, n\}$ and thus $\Sigma, \bigvee_{i=1}^{n} A_i \vdash \Theta$ which is in contradiction to the fact that $\Sigma \mid \bigvee_{i=1}^{n} A_i \mid\!\!\sim_{\Theta} \Delta$ is undefeated. Let $\{A^1, \ldots, A^k\}$ be those members of $\{A_1, \ldots, A_n\}$ for which $\Sigma \mid A^i \mid\!\!\sim_{\Theta} \Delta$ is undefeated and $\{B^1, \ldots, B^j\}$ those members of $\{A_1, \ldots, A_n\}$ for which $\Sigma \mid B^i \mid\!\!\sim_{\Theta} \Delta$ is defeated. Note that $\Sigma, \bigvee_{i=1}^{j} B^i \vdash \Theta$. Thus, $\{A^1, \ldots, A^k\} \subseteq \Xi_{\langle \Sigma, \{\bigvee_{i=1}^{n} A_i\}, \Delta\rangle}$. However, since $\neg A^1, \ldots, \neg A^k, \bigvee_{i=1}^{n} A_i \vdash \bigvee_{i=1}^{j} B^i$, $\neg \Xi_{\langle \Sigma, \{\bigvee_{i=1}^{n} A_i\}, \Delta\rangle}, \Sigma \mid \bigvee_{i=1}^{n} A_i \mid\!\!\sim_{\Theta} \Delta$ is defeated since $\neg \Xi_{\langle \Sigma, \{\bigvee_{i=1}^{n} A_i\}, \Delta\rangle}, \Sigma, \bigvee_{i=1}^{n} A_i \vdash \Theta$. This contradicts our main supposition. ■

Fact 4. If $\Sigma \mid \Gamma \overset{\blacktriangleright}{\underset{\Theta}{\mid\!\!\sim}} \Delta$ is provable, then $\bigwedge \Gamma \dashv\vdash \bigwedge \mathsf{Cn}_{\mathsf{Lit}}(\Gamma)$.

Proof. By Lemma 16, $\Sigma \mid \Gamma \mid\!\!\sim_{\Theta} \Delta$ is provable. By $\wedge\vdash$, $\Sigma \mid \bigwedge \Gamma \mid\!\!\sim_{\Theta} \Delta$ follows. Let $\bigvee_{i=1}^{n} A_i$ where $A_i = \bigwedge \Gamma_i$ be a minimal DNF of $\bigwedge \Gamma$. Then, by applying *cut* to $\bigvee_{i=1}^{n} A_i \vdash \bigwedge \Gamma$ and $\Sigma \mid \bigwedge \Gamma \mid\!\!\sim_{\Theta} \Delta$ we get $\Sigma \mid \bigvee_{i=1}^{n} A_i \mid\!\!\sim_{\Theta} \Delta$. Since $\bigvee_{i=1}^{n} A_i \dashv\vdash \bigwedge \Gamma$, $\Xi_{\langle \Sigma, \Gamma, \Delta\rangle} = \Xi_{\langle \Sigma, \{\bigvee_{i=1}^{n} A_i\}, \Delta\rangle}$. Thus $\Sigma \mid \bigvee_{i=1}^{n} A_i \overset{\blacktriangleright}{\underset{\Theta}{\mid\!\!\sim}} \Delta$ is provable since $\Sigma \mid \Gamma \overset{\blacktriangleright}{\underset{\Theta}{\mid\!\!\sim}} \Delta$ is provable. By Fact 3, $n = 1$. Hence, $\bigwedge \Gamma \dashv\vdash \bigwedge \mathsf{Cn}_{\mathsf{Lit}}(\Gamma)$ since $\mathsf{Cn}_{\mathsf{Lit}}(\Gamma) = \Gamma_1$. ■

Lemma 17. *If* $\Sigma \mid \Gamma \overset{\blacktriangleright}{\underset{\Theta}{\mid\!\!\sim}} \Delta$ *and* $\Sigma \mid \Gamma' \overset{\blacktriangleright}{\underset{\Theta'}{\mid\!\!\sim}} \Delta$ *are provable in* $\mathsf{LEA}_{\mathbb{P}}$ *and* $\Gamma \vdash \bigwedge \Gamma'$ *then* $\Gamma' \vdash \bigwedge \Gamma$.

Proof. Suppose $\Sigma \mid \Gamma \overset{\blacktriangleright}{\underset{\Theta}{\mid\!\!\sim}} \Delta$ and $\Sigma \mid \Gamma' \overset{\blacktriangleright}{\underset{\Theta'}{\mid\!\!\sim}} \Delta$ are provable in $\mathsf{LEA}_{\mathbb{P}}$ and $\Gamma \vdash \bigwedge \Gamma'$. Assume for a contradiction that $\Gamma' \nvdash \bigwedge \Gamma$. By Fact 4, $\bigwedge \Gamma \dashv\vdash \bigwedge \mathsf{Cn}_{\mathsf{Lit}}(\Gamma)$ and $\bigwedge \Gamma' \dashv\vdash \bigwedge \mathsf{Cn}_{\mathsf{Lit}}(\Gamma')$. Thus, $\Gamma \vdash \mathsf{Cn}_{\mathsf{Lit}}(\Gamma')$ and $\mathsf{Cn}_{\mathsf{Lit}}(\Gamma) \vdash \mathsf{Cn}_{\mathsf{Lit}}(\Gamma')$, but $\mathsf{Cn}_{\mathsf{Lit}}(\Gamma') \nvdash \mathsf{Cn}_{\mathsf{Lit}}(\Gamma)$. Since $\Sigma \mid \Gamma' \overset{\blacktriangleright}{\underset{\Theta'}{\mid\!\!\sim}} \Delta$ is provable, by Lemma 16, $\Sigma \mid \Gamma' \mid\!\!\sim_{\Theta} \Delta$ is also provable. By *cut*, $\Sigma \mid \mathsf{Cn}_{\mathsf{Lit}}(\Gamma') \mid\!\!\sim_{\Theta'} \Delta$ is obtained. Thus, $\mathsf{Cn}_{\mathsf{Lit}}(\Gamma') \subseteq \Xi_{\langle \Sigma, \Gamma, \Delta\rangle}$. But then $\neg \Xi_{\langle \Sigma, \Gamma, \Delta\rangle}, \Sigma \mid \Gamma \mid\!\!\sim_{\Theta} \Delta$ is defeated since $\neg \Xi_{\langle \Sigma, \Gamma, \Delta\rangle} \cup \Gamma$ is inconsistent. This contradicts our supposition. ■

We can also show that $\mathsf{LEA}_{\mathbb{P}}$ is congruential—i.e. that sequents with logically equivalent antecedents/succedents are co-provable against identical background and defeater sets. In formulating this claim, we let $\mid\!\frac{[\blacktriangleright]}{\Theta}$ range over sequents constructed with either type of turnstile.

Lemma 18. *If $\bigwedge \Gamma \dashv\vdash \bigwedge \Gamma'$ and $\bigvee \Delta \dashv\vdash \bigvee \Delta$, then $\Sigma \mid \Gamma \mid\!\frac{[\blacktriangleright]}{\Theta} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$ iff $\Sigma \mid \Gamma' \mid\!\frac{[\blacktriangleright]}{\Theta} \Delta'$ is.*

Proof. Lemma 7 gives us the result for all sequents whose proof does not involve STR. To show that the lemma holds for $\blacktriangleright$-sequents we consider cases of logically equivalent antecedents and succedents separately. From Lemma 17, it follows that the lemma holds for $\blacktriangleright$-sequents with logically equivalent antecedents. From Lemma 7 and Definition 19, it follows that premises of STR with logically equivalent succedents will have identical background sets, thus ensuring their co-provability. Since the rules for logical connectives in $\mathsf{LE}^{\blacktriangleright}$ are admissible in $\mathsf{LEA}_{\mathbb{P}}$, this is sufficient to guarantee the result holds for all $\blacktriangleright$-sequents. Finally, for sequents proved via ABD, we obtain our result in the manner of Lemma 7's proof, namely, by applying *cut* to LK-provable sequents with empty defeater sets. ■

We can now see that the $\blacktriangleright$-sequents provable in $\mathsf{LEA}_{\mathbb{P}}$ exhibit all of the properties associated with explanatory arguments.

Theorem 19. *The $\blacktriangleright$-sequents that are theorems of $\mathsf{LEA}_{\mathbb{P}}$ are (1) defeasible, (2) premise-consistent, (3) material, (4) minimal, (5) sturdy, (6) detachable, and (7) congruential.*

Proof. (1) Defeasibility (Definition 1) follows from Corollary 12.
(2) Premise-consistency (Definition 3) follows from Corollary 13.
(3) Materiality (Definition 4) follows from Corollary 14.
(4) Minimality (Definition 2) follows from Lemma 17.
(5) Sturdiness follows from STR.
(6) Detachability follows from ABD.
(7) Congruentiality (Definition 6) follows from Lemma 18

■

Lastly, we show that explanatory arguments that correspond to $\blacktriangleright$-sequents are indeed those of *immediate* explanations in virtue of the fact that their consequence relations are not transitive.

Proposition 20. It is possible that $\Sigma \mid \Gamma \mid\!\frac{\blacktriangleright}{\Theta} A, \Delta$ and $\Sigma' \mid \Gamma', A \mid\!\frac{\blacktriangleright}{\Psi} \Delta'$ are provable in $\mathsf{LEA}_{\mathbb{P}}$, but $\Sigma', \Sigma \mid \Gamma', \Gamma \mid\!\frac{\blacktriangleright}{\Theta \cup \Psi} \Delta, \Delta'$ is not.

Proof. Let $\mathbb{P} = \{\langle\{b\}, \{p, p'\}, \{d\}, \{s\}\rangle, \langle\{b\}, \{q, s\}, \{d\}, \{t\}\rangle, \langle\{b\}, \{p', q\}, \{d\}, \{t\}\rangle\}$. Then $b \mid p', p \mid\!\frac{\blacktriangleright}{\{d,s\}} s$; $b \mid q, s \mid\!\frac{\blacktriangleright}{\{d,t\}} t$; and $b \mid p', q \mid\!\frac{\blacktriangleright}{\{d,t\}} t$ are provable in $\mathsf{LEA}_{\mathbb{P}}$, but $b \mid p', p, q \mid\!\frac{\blacktriangleright}{\{d,t\}} t$ is not since $\{p', p, q\}$ is not a minimal explanation in view of $b \mid p', q \mid\!\frac{\blacktriangleright}{\{d,t\}} t$ (see Lemma 17). ■

4.3.3 Discussion

While $\mathsf{LEA}_{\mathbb{P}}$ gives us the six properties we've associated with best explanations, there are outstanding issues with the calculus which deserve treatment. Thus, before proceeding to the extension of $\mathsf{LEA}_{\mathbb{P}}$, we would like to highlight some of its limitations.

First, as we've mentioned, proofs in $\mathsf{LEA}_{\mathbb{P}}$ contain applications of *cut* that were not available in $\mathsf{SC}_{\mathbb{P}}^{\boldsymbol{\Theta}}$. Since our rule for ABD aims to capture abduction's ampliative character, the tree segments that occur below the use of this rule behave as if a new axiom were introduced. Consequently, applications of the cut rule can only be permuted up to the abductive step, at which point, a cut-free proof may not be available.

Proposition 21. There are sequents provable in $\mathsf{LEA}_{\mathbb{P}}$ that do not have cut-free proofs.

Proof. The following example will suffice.

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ r_1 \mid p \vdash^{\blacktriangleright}_{\{\neg p,r_2\}} r_2\end{array}\qquad \begin{array}{c}\mathcal{D}_2\\ \vdots\\ r_1 \mid q \vdash_{\{\neg p,r_2\}} r_2\end{array}}{r_1 \mid q \vdash_{\{\neg p,r_2\}} p}\,\mathsf{ABD}\qquad \begin{array}{c}\mathcal{D}_3\\ \vdots\\ r_1 \mid p \vdash_{\{r_2\}} r_3\end{array}}{r_1 \mid q \vdash_{\{\neg p,r_2\}} r_3}\,cut$$

Under the assumption that $\langle\{r_1\},\{q\},\Theta,\{r_3\}\rangle \notin \mathbb{P}$ such that $\Theta \subseteq \{\neg p, r_2\}$, there is no proof of the end-sequent that does not use *cut*. ■

A second issue with the calculus arises from our decision to work with a non-exclusive conception of *best explains*. This conception leaves open the possibility that incompatible hypotheses are both deemed 'the best.' In an axiomatized system based on classical logic and equipped with a rule for IBE, this means that it is possible to derive contradictions from incompatible abducibles. Here is an example of how this might occur in $\mathsf{LEA}_{\mathbb{P}}$.

Example 11.

$$\mathcal{D}_1 = \dfrac{\dfrac{}{r \mid p \vdash_{\{s\}} q}\,prop.ax.}{r \mid p \vdash_{\{s,q\}} q}\,\mathsf{DE}\qquad \mathcal{D}_2 = \dfrac{\dfrac{}{r \mid t \vdash_{\{v\}} q}\,prop.ax.}{r \mid t \vdash_{\{v,q\}} q}\,\mathsf{DE}\qquad \mathcal{D}_3 = \dfrac{\dfrac{}{r \mid \neg p \vdash_{\{u\}} q}\,prop.ax.}{r \mid \neg p \vdash_{\{u,q\}} q}\,\mathsf{DE}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\mathcal{D}_1}{\neg\neg p, \neg t, r \mid p \vdash_{\{s,q\}} q}\,\mathsf{BE}\qquad \mathcal{D}_1}{r \mid p \vdash^{\blacktriangleright}_{\{s,q\}} q}\,\mathsf{STR}\qquad \mathcal{D}_2}{r \mid t \vdash_{\{s,v,q\}} p}\,\mathsf{ABD}\qquad \dfrac{\dfrac{\dfrac{\mathcal{D}_3}{\neg p, \neg t, r \mid \neg p \vdash_{\{u,q\}} q}\,\mathsf{BE}\qquad \mathcal{D}_3}{r \mid \neg p \vdash^{\blacktriangleright}_{\{u,q\}} q}\,\mathsf{STR}\qquad \mathcal{D}_2}{r \mid t \vdash_{\{u,v,q\}} \neg p}\,\mathsf{ABD}}{r \mid t \vdash_{\{s,v,u,q\}} p \wedge \neg p}\,\vdash\wedge$$

In this example, the hypotheses p and $\neg p$ both pass the sturdiness test and therefore 'tie' for the title of *best explanation*. After abducing both hypotheses, one can derive a contradiction. The threat of explosion posed by such scenarios is clearly undesirable. In earlier versions of the system, we considered forcing the abduced formula into the defeater set of ABD's conclusion, thus ensuring defeat when contradictory abducibles are conjoined. But while such an alteration solves situations such as the one in Example 11, it does not solve similar ones in which abducibles are are *materially incompatible*, rather than logically inconsistent—e.g. when p and q are abducible, but $\Sigma \mid p \vdash_{\Theta} \neg q$ is a proper axiom. Resolving the difficulties posed by these scenarios would require a far more dramatic change to the calculus.

While the persistence of this problem—what we'll call *post-abductive explosion*—is unfortunate, there are several mollifying considerations. First, unless explicit constraints are imposed on proper axioms, the threat of explosion looms over any axiomatized system based on classical logic. For example, in ours, there is nothing to prevent $\langle\{r\},\{p\},\{s\},\{q\}\rangle, \langle\{r\},\{t\},\{v\},\{\neg q\}\rangle \in \mathbb{P}$, such that one may derive $r \mid t, p \vdash_{\{s,v\}} q \wedge \neg q$. Second, post-abductive explosion may be thought to reveal an important and natural consequence of the non-exclusive conception of *best explanation*, and therefore its presence in our system is simply a product of the latter's accurate formalization of that conception. Third, the problem only arises when the system's initial information state (i.e. the set of proper axioms) is, to a certain extent, incoherent. Indeed, it is almost paradoxical to think that a hypothesis and its denial could both be potential explanations of the same phenomenon, since doing so runs contrary to the common assumption that explanantia 'make a difference' to their explananda. In this sense, post-abductive explosion signals to the reasoner that she ought to 'clean up' her initial assumptions before drawing explanatory and abductive conclusions.[16] A practitioner who remains troubled by post-abductive explosion could, of course, insist that STR and ABD only occur once per *derivation* rather than once per branch, as we currently have it. Imposing this restriction would prevent explosion.

4.4 *The extension* $\mathsf{LEA}^+_{\mathbb{P}}$

We shall now demonstrate how the system $\mathsf{LEA}_{\mathbb{P}}$ and the language $\mathcal{L}$ over which it is defined may be extended to include an object-language expression for *best explains why.* We begin with the syntax of the extended language $\mathcal{L}_{\blacktriangleright}$.

Definition 22 (Syntax of $\mathcal{L}_{\blacktriangleright}$)**.**

$$\text{(1) If } A \in \mathcal{L} \text{ then } A \in \mathcal{L}_{\blacktriangleright}.$$
$$\text{(2) If } A, B \in \mathcal{L} \text{ then } A \blacktriangleright B \in \mathcal{L}_{\blacktriangleright}$$

The expression '$A \blacktriangleright B$' is intended to be read as '*A best explains why B*.' Note that the syntactic definition for $\blacktriangleright$ is not recursive with respect to $\mathcal{L}_{\blacktriangleright}$. Consequently, the operator $\blacktriangleright$ is non-iterative. We impose this syntactic constraint on the grounds that, as argued in Section 4, the "...best explains why..." locution does not appear to iterate in natural languages—at least not in English.

The rules in Figure 3 define the calculus $\mathsf{LEA}^+_{\mathbb{P}}$ over $\mathcal{L}_{\blacktriangleright}$. While the rules of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ apply to all the formulas in $\mathcal{L}_{\blacktriangleright}$, the rules for $\mathsf{LE}^{\blacktriangleright}$ are restricted to the fragment $\mathcal{L} \cap \mathcal{L}_{\blacktriangleright}$. This restriction is intended to prevent the generation of ill-formed formulas along a derivation, e.g. $A \blacktriangleright (B \blacktriangleright C)$.

The rules of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$ apply to all the formulas in $\mathcal{L}_{\blacktriangleright}$.

The rules of $\mathsf{LE}^{\blacktriangleright}$ apply to all formulas in the fragment $\mathcal{L} \cap \mathcal{L}_{\blacktriangleright}$.

Rules for $\blacktriangleright$

$$\frac{\Sigma \mid A \mathrel{|\!\!\!\frac{\;\blacktriangleright\;}{\Theta}} B \qquad \Sigma' \mid \Gamma \mathrel{|\!\!\!\frac{\;\;}{\Psi \cup \{B\}}} B}{\Sigma', \Sigma \mid \Gamma, A \blacktriangleright B \mathrel{|\!\!\!\frac{\;\;}{\Theta \cup \Psi \cup \{B\}}} A} \;\blacktriangleright\!\vdash^{\ddagger\ddagger} \qquad\qquad \frac{\Sigma \mid A \mathrel{|\!\!\!\frac{\;\blacktriangleright\;}{\Theta}} B}{\Sigma \mid \cdot \mathrel{|\!\!\!\frac{\;\;}{\Theta}} A \blacktriangleright B} \;\vdash\!\blacktriangleright$$

‡‡ Provided that $A \nvdash \Gamma$.

Figure 3.: Rules of $\mathsf{LEA}^+_{\mathbb{P}}$

[16] Of course, scientists do find themselves confronting inconsistent evidence and assumptions of the sort that might give rise to post-abductive explosion, and it is desirable to have a system that handles such situations. To that end, we would like to explore alternative constructions using other non-classical logics, such as those formulated to capture paraconsistency.

As promised, the extension $\mathsf{LEA}^+_{\mathbb{P}}$ provides introduction (right) and elimination (left) rules for the *best-explains-why* operator, i.e. $\blacktriangleright$. Unlike the logical rules in $\mathsf{LEA}_{\mathbb{P}}$, these rules apply irrespective of whether the active formulas' main operators are negation.

The $\blacktriangleright\vdash$ rule ought to be familiar—it is essentially ABD with the abducible formula joined to the succedent of the premises by the $\blacktriangleright$-connective and appended to the antecedent of the conclusion. In fact, it is readily observable that $\blacktriangleright\vdash$ is derivable from ABD and LW.

Lemma 22. *$\blacktriangleright\vdash$ is derivable in $\mathsf{LEA}^+_{\mathbb{P}}$.*

Proof. From the premises of $\blacktriangleright\vdash$ apply ABD and weaken the resulting antecedent by $A \blacktriangleright B$. ∎

With the formulation of $\blacktriangleright\vdash$, we have achieved the main goal of this paper: to formalize IBE. As a rule of inference, IBE is represented by the defeasible sequent in the conclusion. In keeping with our preferred reading of the proof rules, $\blacktriangleright\vdash$ represents a meta-rule of inference; it tells us what other inferences need to be available to reasoners in order for them to legitimately deploy IBE. Let us read this meta-rule. $\blacktriangleright\vdash$ says that anyone who is entitled to the best explanatory argument (A) for B and who has independent evidence (Γ) of B is entitled to conclude A on the basis of that evidence (Γ) and the *fact* that A *best explains* B. Roughly put, the end-sequent tells us that if A best explains B and if we have evidence that B is the case, then we may conclude that A is also the case. The background sets in this rule inform us that we are entitled to draw an inference to the best explanation so long as doing so is consistent with (i.e. not defeated by) the background information that informs both the explanatory argument from A to B as well as the evidential inference to B.

In contrast with the turnstile-level rule of ABD, $\blacktriangleright\vdash$ requires its premises to have single-formula antecedents and succedents. Part of the motivation for this restriction comes from our desire to present the *strongest* version of IBE that we can. But another source of motivation lies in the fact that our target is the *exhaustive* explanatory relation. If A were only one of many formulas that together exhaustively explain, then $A \blacktriangleright B$ would only capture a partial explanatory relation. To better represent *exhaustive* explanations, we thus restrict the first premise to single-antecedent sequents.

Yet further motivation is to be found in the observation that if side formulas were permitted in the succedents of the premises, it would be unclear whether and where those formulas ought to occur in the conclusion. Should they disappear as the entire succedent does in ABD, or should they carry down into the succedent of the conclusion? In the latter case we would be left with a much weaker inference rule than the one we have currently. If we took the former approach, we would have to account for the fact that a reasoner is entitled to assume that a particular formula from the succedent is explained when the explanatory argument only tells us that at least one of those formulas is explained. Given these undesirable consequences, we have chosen to formulate the rule with single-succedent premises.

We now proceed to consider $\vdash\blacktriangleright$. While this rule resembles the right rule for the material conditional ($\vdash\supset$) in LK, since $\mathsf{LEA}^+_{\mathbb{P}}$ lacks the LK rules for negation, it has a very different meaning. It differs from $\vdash\supset$ in two other respects. First, the premise contains no sides formulas. Like considerations that motivate single-succedents in $\blacktriangleright\vdash$, the presence of side formulas in the succedents of $\vdash\blacktriangleright$ would raise questions about how entitlement to an explanatory claim about a particular formula is obtained on the basis of explanatory arguments for (only) some member of a set. Second, while the premise is a $\blacktriangleright$-sequent, the conclusion is not. This feature reflects the sense in which formulas whose main operator is $\blacktriangleright$ are *explicitly* explanatory claims. As such, they can (and must!) enter into reasoning patterns that do not consist in the making of explanatory arguments, and thus they belong to the class of sequents whose turnstile is unadorned by $\blacktriangleright$.

We are now in a position to demonstrate that the $\blacktriangleright$-connective encodes the consequence relation, $\vdash^{\blacktriangleright}_{\Theta}$. Since $\mathsf{LEA}^{+}_{\mathbb{P}}$ invokes two distinct classes of consequence relations, we will not be able to state a traditional deduction theorem. Instead, we have the claim that an expression used in one logic (that of the unadorned turnstile, $\vdash_{\Theta}$) encodes the rules of another logic ($\vdash^{\blacktriangleright}_{\Theta}$). To distinguish this claim from the standard deduction theorem, we refer to it as a *quasi-deduction theorem.*

Theorem 23 (Quasi-Deduction Theorem)**.** *$\Sigma \mid A \vdash^{\blacktriangleright}_{\Theta} B$ is provable in $\mathsf{LEA}^{+}_{\mathbb{P}}$ iff $\Sigma \mid \cdot \vdash_{\Theta} A \blacktriangleright B$ is.*

Proof. ($\Rightarrow$) Follows from $\vdash\blacktriangleright$.

($\Leftarrow$) Since there are no side formulas in $\vdash\blacktriangleright$, $A \blacktriangleright B$ must be principal in the last rule and $\Sigma \mid A \vdash^{\blacktriangleright}_{\Theta} B$ must be its premise. ■

We can now show that $\mathsf{LEA}^{+}_{\mathbb{P}}$ is a conservative extension of $\mathsf{LEA}_{\mathbb{P}}$. As with Lemma 18, we let $\vdash^{[\blacktriangleright]}_{\Theta}$ range over sequents constructed with either type of turnstile.

Theorem 24. *If $\Sigma \mid \Gamma \vdash^{[\blacktriangleright]}_{\Theta} \Delta$ is provable in $\mathsf{LEA}^{+}_{\mathbb{P}}$ and $\Sigma, \Theta, \Gamma, \Delta \subset \mathcal{L}$, then $\Sigma \mid \Gamma \vdash^{[\blacktriangleright]}_{\Theta} \Delta$ is provable in $\mathsf{LEA}_{\mathbb{P}}$.*

Proof. From Definition 20 and the rules of $\mathsf{LEA}^{+}_{\mathbb{P}}$, it follows that any sequent that is provable in $\mathsf{LEA}^{+}_{\mathbb{P}}$ without the use of $\blacktriangleright\vdash$ or $\vdash\blacktriangleright$ is provable in $\mathsf{LEA}_{\mathbb{P}}$. That leaves us with sequents whose proof involves the application of *cut* to a $\blacktriangleright$-formula. To satisfy our hypothesis, the cut formula in such proofs must be principal in the last rule. We therefore provide a parallel reduction for $\blacktriangleright\vdash\backslash\vdash\blacktriangleright$, as follows.

$$\dfrac{\dfrac{\overset{\mathcal{D}_1}{\vdots} \\ \Sigma \mid A \vdash^{\blacktriangleright}_{\Theta} B}{\Sigma \mid \cdot \vdash_{\Theta} A \blacktriangleright B}\;{\scriptstyle \vdash\blacktriangleright} \qquad \dfrac{\overset{\mathcal{D}_2}{\vdots} \\ \Sigma' \mid A \vdash^{\blacktriangleright}_{\Theta'} B \qquad \overset{\mathcal{D}_3}{\vdots} \\ \Sigma'' \mid \Gamma \vdash_{\Psi} B}{\Sigma'', \Sigma' \mid \Gamma, A \blacktriangleright B \vdash_{\Theta' \cup \Psi} A}\;{\scriptstyle \blacktriangleright\vdash}}{\Sigma'', \Sigma', \Sigma \mid \Gamma \vdash_{\Theta \cup \Theta' \cup \Psi} A}\;cut$$

$$\rightsquigarrow$$

$$\dfrac{\overset{\mathcal{D}_2}{\vdots} \\ \Sigma' \mid A \vdash^{\blacktriangleright}_{\Theta'} B \qquad \dfrac{\dfrac{\overset{\mathcal{D}_3}{\vdots} \\ \Sigma'' \mid \Gamma \vdash_{\Psi} B}{\Sigma'', \Sigma', \Sigma \mid \Gamma \vdash_{\Psi} B}\;\mathsf{BE}}{\Sigma'', \Sigma', \Sigma \mid \Gamma \vdash_{\Theta \cup \Psi} B}\;\mathsf{DE}}{\Sigma'', \Sigma', \Sigma \mid \Gamma \vdash_{\Theta \cup \Theta' \cup \Psi} A}\;\mathsf{ABD}$$

■

5. Conclusion

In this paper we have sought to formalize the qualitative treatment of IBE in a manner that does justice to the distinctly explanatory character of its premises. To that end, we have formulated IBE as the elimination rule for an explicitly explanatory connective, $\blacktriangleright$, which is intended to encode (at least part of) the meaning associated with the English expression "That... best explains why..." Through the proof of a quasi-deduction theorem, we have shown that this connective makes explicit an inference rule for explanatory arguments that is represented in our meta-language as a class of defeasible consequence relations, $\mid\!\overset{\blacktriangleright}{\sim}_{\Theta}$. Finally, we have demonstrated that provable $\blacktriangleright$-sequents exhibit all of the basic features associated with our best explanations (Theorem 19). This result obtains, in large part, because the system contains a rule for the test of sturdiness, i.e. STR. Since one of the central tenets of DIME is that explananda are the sturdy consequences of our best explanantia, these formal results stand as a testament to the power and precision of that model.

As we mentioned in section 3, we have sought to restrict explanatory claims and arguments to those that are irreflexive. On closer inspection, however, we found that to do so in a classical sequent system requires the attainment of a stronger condition, namely, classical non-validity, or what we've termed *materiality*, and so we have provided calculi that realize this property. For various reasons, one might wish to relax this constraint. To do so, one simply removes the reference to succedents in the defeater sets for rules where their addition is explicit, namely, $\mathsf{STR}, \mathsf{ABD}, \blacktriangleright\vdash$, and $\vdash\blacktriangleright$. Note that STR requires the defeater set of the premise that proceeds addition of the succedent to be nonempty. So, even without adding the succedent, STR will still yield defeasible and premise-consistent conclusions. One might also wish to allow competitors in the sturdiness test to be valid, in which case, one would simply remove Δ from the defeater set in the definition of Antecedent Set, i.e. $\mathbf{S}_{\langle\Sigma,\Delta\rangle}$ (Definition 14). The fact that removing the materiality constraint from the calculus is so easy speaks to the flexibility of treating classes of defeasible consequence relations as the basis for an account of explanatory arguments.

There are several avenues of development we think would be worth pursuing. For instance, we speculate that a version of $\mathsf{LEA}^{+}_{\mathbb{P}}$ formulated on the basis of a logic for paraconsistency might overcome the problem of post-abductive explosion. The addition of rules for modal operators would also be desirable given the important relationship between explanation and counterfactuals. Finally, a variant that incorporated *hyperintensional* conceptions of explanation, irreflexivity, or defeat might overcome some of the limitations of our calculi. We conclude, therefore, with an invitation to further investigate logics for best explanations.

References

Aliseda, A. (2006). *Abductive Reasoning: Logical Investigations into Discovery and Explanation.* Springer Netherlands.

Avron, A. (2016). RM and its Nice Properties. In K. Bimbó (Ed.), *J. Michael Dunn on Information Based Logic*, 15–43.

Beirlaen, M., & Aliseda, A. (2014). A Conditional Logic for Abduction. *Synthese*, *191*(15), 3733–3758.

Bird, A. (2007). Inference to the Only Explanation. *Philosophy and Phenomenological Research*, *74*(2), 424–432.

Boutilier, C., & Becher, V. (1995). Abduction as Belief Revision. *Artificial Intelligence*, *77*(1), 43 - 94.

Campos, D. G. (2011). On the Distinction Between Peirce's Abduction and Lipton's Inference to the Best Explanation. *Synthese*, *180*(3), 419–442.

Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science.* Cambridge, MA: Cambridge University Press.

D'Agostino, M., & Mondadori, M. (1994). The Taming of the Cut. Classical Refutations with Analytic Cut. *Journal of Logic and Computation*, *4*(3), 285-319.
Day, T., & Kincaid, H. (1994). Putting Inference to the Best Explanation in its Place. *Synthese*, *98*(2), 271–295.
Douven, I. (1999). Inference to the Best Explanation Made Coherent. *Philosophy of Science*, *66*, 424-S435.
Douven, I. (2002). Testing Inference to the Best Explanation. *Synthese*, *130*(3), 355–377.
Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press.
Fine, K. (2012). The Pure Logic of Ground. *Review of Symbolic Logic*, *5*(1), 1–25.
Flach, P. (1996a). Abduction and Induction: Syllogistic and Inferential Perspectives. In *Abductive and Inductive Reasoning Workshop Notes*, 31–35.
Flach, P. (1996b). Rationality Postulates for Induction. In Y. Shoham (Ed.), *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 1996)*, 267–281. De Zeeuwse Stromen.
Fodor, J. A. (1974). Special Sciences (Or: The Disunity of Science as a Working Hypothesis). *Synthese*, *28*(2), 97–115.
Friedman, M. (1974). Explanation and Scientific Understanding. *Journal of Philosophy*, *71*(1), 5–19.
Gabbay, D., & Woods, J. (2006). Advice on Abductive Logic. *Logic Journal of IGPL*, *14*(2), 189–219.
Glass, D. (2007). Coherence Measures and Inference to the Best Explanation. *Synthese*, *157*(3), 275–296.
Harman, G. (1965). The Inference to the Best Explanation. *Philosophical review*, *74*, 88–95.
Halonen, I., & Hintikka, J. (2005, Jan 01). Toward a Theory of the Process of Explanation. *Synthese*, *143*(1), 5–61.
Hempel, C. G. (1965). *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science.* New York: Free Press.
Hempel, C. G. (1966). *Philosophy of Natural Science.* Englewood Cliffs, N.J.: Prentice-Hall.
Jenkins, C. S. (2011). Is Metaphysical Dependence Irreflexive? *The Monist*, *94*(2), 267–276.
Khalifa, K., Millson, J., & Risjord, M. (2018a). IBE: Fundamentalism's Failures. In K. McCain & T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation.* Oxford: Oxford University Press.
Khalifa, K., Millson, J., & Risjord, M. (2018b). Inference, Explanation, and Asymmetry. *Synthese*.
Kitcher, P. (1989). Explanatory Unification and the Causal Structure of the World. In P. Kitcher & W. C. Salmon (Eds.), *Scientific Explanation.* Minneapolis: University of Minnesota Press, 410–506.
Kyburg, H. E. (1965). Salmon's Paper. *Philosophy of Science*, *32*(2), 147–151.
Lange, M. (2009). Why Do the Laws Explain Why? In T. Handfield (Ed.), *Dispositions and Causes.* Oxford University Press.
Lipton, P. (2004). *Inference to the Best Explanation.* Routledge.
Litland, J. E. (2016). Pure Logic of Many-Many Ground. *Journal of Philosophical Logic*, *45*(5), 531–577.
Lobo, J., & Uzcátegui, C. (1997). Abductive Consequence Relations. *Artificial Intelligence*, *89*(1), 149–171.
Magnani, L. (2001). *Abduction, Reason, and Science : Processes of Discovery and Explanation.* New York: Kluwer Academic/Plenum Publishers.

Magnani, L., Carnielli, W., & Pizzi, C. (eds.) (2012). Special Issue: Formal Representations in Model-Based Reasoning and Abduction. *Logic Journal of the IGPL*, *20*(1), 367–532.

Mayer, M. C., & Pirri, F. (1993). First Order Abduction via Tableau and Sequent Calculi. *Logic Journal of IGPL*, *1*(1), 99–117.

Mayer, M. C., & Pirri, F. (1996). Abduction is not Deduction-in-Reverse. *Journal of the IGPL*, *4*(1), 95-108.

Meheus, J. (2005). Empirical Progress and Ampliative Adaptive Logics. *Poznan Studies in the Philosophy of the Sciences and the Humanities*, *83*(1), 193–217.

Meheus, J., & Batens, D. (2006). A Formal Logic for Abductive Reasoning. *Logic Journal of the IGPL*, *14*(2), 221–236.

Meheus, J., Verhoeven, L., Van Dyck, M., & Provijn, D. (2002). Ampliative Adaptive Logics and the Foundation of Logic-based Approaches to Abduction. In *Logical and Computational Aspects of Model-based Reasoning* (pp. 39–71). Springer.

Millson, J. (2018). A Cut-free Sequent Calculus for Defeasible Erotetic Inferences. *Studia Logica*.

Millson, J., Khalifa, K., & Risjord, M. (2018). Inferentialist-Expressivism for Explanatory Vocabulary. In L. K. Ondřej Beran Vojtěch Kolman (Ed.), *From Rules to Meaning: New Essays in Inferentialism.* Routledge.

Mitchell, S. D. (2003). *Biological Complexity and Integrative Pluralism.* Cambridge University Press.

Piazza, M., & Pulcini, G. (2016). Uniqueness of Axiomatic Extensions of Cut-free Classical Propositional Logic. *Logic Journal of IGPL*, *24*(5), 708-718.

Piazza, M., & Pulcini, G. (2017). Unifying Logics via Context-sensitiveness. *Journal of Logic and Computation*, *27*(1), 21.

Pino-Pérez, R., & Uzcátegui, C. (1999). Jumping to Explanations versus Jumping to Conclusions. *Artificial Intelligence*, *111*(1), 131–169.

Pino-Pérez, R., & Uzcátegui, C. (2000). Ordering Explanations and the Structural Rules for Abduction. In B. Selman, A. Cohn, & F. Giunchiglia (Eds.), *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR-2000)*, 637–646.

Pino-Pérez, R., & Uzcátegui, C. (2003). Preferences and Explanations. *Artificial Intelligence*, *149*(1), 1–30.

Poggiolesi, F. (2018). On Constructing a Logic for the Notion of Complete and Immediate Formal Grounding. *Synthese*, *195*(3), 1231–1254.

Pollock, J. (2015). *Knowledge and Justification.* Princeton University Press.

Psillos, S. (2000). Abduction: Between Conceptual Richness and Computational Complexity. In A. Kakas & P. Flach (Eds.), *Abduction and Induction: Essays in their Relation and Integration* (Vol. 18). Dordrecht: Kluwer, 605-625.

Psillos, S. (2002). Simply the Best: A Case for Abduction. In A. Kakas & F. Sadri (Eds.), *Computational Logic: Logic Programming and Beyond : Essays in Honour of Robert A. Kowalski, Part II.* Berlin: Springer, 83–93.

Salmon, W. C., & Jeffrey, R. (1971). *Statistical Explanation & Statistical Relevance.* University of Pittsburgh Press.

Schnieder, B. (2011). A Logic for 'Because'. *Review of Symbolic Logic*, *4*(3), 445–465.

Schupbach, J. N., & Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science*, *78*(1), 105–127.

Schurz, G. (2008). Patterns of Abduction. *Synthese*, *164*(2), 201–234.

Schurz, G., & Lambert, K. (1994). Outline of a Theory of Scientific Understanding. *Synthese*, *101*(1), 65–120.

Skow, B. (2016). *Reasons Why.* Oxford University Press UK.

Skyrms, B. (1980). *Causal Necessity: A Pragmatic Investigation of the Necessity of*

Laws. New Haven, CT: Yale University Press.
Smullyan, R. (1968). *First-order logic*. Springer-Verlag.
Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese*, *167*(1), 125–143.
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.

Appendix A. Proof of Cut-Elimination Theorem for $\mathsf{SC}_{\mathbb{P}}^{\Theta}$ (Lemma 5)

Notation

We will sometimes make use of Smullyan's distinction between α and β formulas (see (Smullyan, 1968, p. 21)) to reduce the need for case distinctions in the proofs, since this allows me to abstract slightly. We use the following table of α and β formulas (where l is a literal):

α	α_1	α_2	β_1	β_2	β
$A \wedge B$	A	B	$\neg A$	$\neg B$	$\neg(A \wedge B)$
$\neg(A \vee B)$	$\neg A$	$\neg B$	A	B	$A \vee B$
$\neg\neg A$	A		$\neg A$		$\neg A$
l	l		$-l$		$-l$

The first line of the table reads, for instance, as follows: if $C = \alpha = A \wedge B$ then $\alpha_1 = A$, $\alpha_2 = B$, $\beta_1 = \neg A$, $\beta_2 = \neg B$ and $\beta = \neg(A \wedge B)$, if $C = \beta = \neg(A \wedge B)$ then $\alpha = A \wedge B$, $\alpha_1 = A$, $\alpha_2 = B$, $\beta_1 = \neg A$, and $\beta_2 = \neg B$.

We also consider α- and corresponding β-rules:

α-Rule	β-Rule
$\wedge\vdash$	$\vdash\neg\wedge$
$\neg\neg\vdash$	$\vdash\neg\neg$
$\neg\vee\vdash$	$\vdash\vee$
$\vdash\wedge$	$\neg\vee\vdash$

We write $\Sigma \mid \Gamma \models_{\Theta} \Delta$ if there is a *cut*-free proof of $\Sigma \mid \Gamma \vdash_{\Theta} \Delta$ in $\mathsf{SC}_{\mathbb{P}}^{\Theta}$.

Outline of the proof

We use the following strategy. We prove properties 1 and 2 below (properties 3 and 4 are auxiliary for this purpose):

(1) Show that if $\Sigma \mid \Gamma, A \models_{\Theta} \Delta$ then $\Sigma \mid \Gamma \models_{\Theta} \neg A, \Delta$. In signs:

$$\frac{\Sigma \mid \Gamma, A \models_{\Theta} \Delta}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \neg A} \models\neg$$

(2) Show that if $\Sigma \mid \Gamma \models_{\Theta} \Delta, A$; $\Sigma' \mid \Gamma' \models_{\Theta'} \Delta', \neg A$; and $\Sigma, \Sigma', \Gamma, \Gamma' \nvdash \Theta, \Theta'$, then

$\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{\equiv}} \Delta, \Delta'$. In signs:

$$\frac{\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, A \quad \Sigma' \mid \Gamma' \mathrel{|\!\!\underset{\Theta'}{\equiv}} \Delta', \neg A}{\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{\equiv}} \Delta, \Delta'} \models\top$$

(3) If $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \alpha$ then $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \alpha_1$, $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \alpha_2$, $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \alpha_1, \beta_2$, and $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \alpha_2, \beta_1$.

(4) If $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \beta$ then $\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \beta_1, \beta_2$.

With properties 1 and 2 we get the admissibility of cut as follows. Suppose we have given the following proof:

$$\frac{\begin{array}{c}\mathcal{D}\\ \vdots \\ \Sigma \mid \Gamma, A \mathrel{|\!\!\underset{\Theta}{-}} \Delta\end{array} \quad \begin{array}{c}\mathcal{D}'\\ \vdots \\ \Sigma' \mid \Gamma' \mathrel{|\!\!\underset{\Theta'}{-}} \Delta', A\end{array}}{\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{-}} \Delta, \Delta'}\, cut$$

It reduces to:

$$\frac{\dfrac{\Sigma \mid \Gamma, A \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta}{\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \neg A} \models\neg \qquad \Sigma' \mid \Gamma' \mathrel{|\!\!\underset{\Theta'}{\equiv}} \Delta', A}{\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{\equiv}} \Delta, \Delta'} \models\top$$

We now show properties 1 and 2.

Property 1

We show the following:

$$\frac{\Sigma \mid \Gamma, \alpha \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta}{\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{\equiv}} \Delta, \beta} \models\neg$$

We show this via a nested induction over the length of α (outer induction) and the length of the underlying proof $\mathcal{D}$ of $\Sigma \mid \Gamma, \alpha \mathrel{|\!\!\underset{\Theta}{-}} \Delta$ (inner induction).

We first cover generally the case that the length of the proof is one and α is of arbitrary length. Suppose $\Sigma \mid \Gamma, \alpha \mathrel{|\!\!\underset{\Theta}{-}} \Delta$ is a proper axiom. Then the statement follows since $\mathbb{P}$ is closed under contraposition. Suppose $\emptyset \mid \Gamma, \alpha \mathrel{|\!\!\underset{\emptyset}{}} \Delta$ is a logical axiom, then the statement follows by the adequacy of Smullyan's calculus for classical logic.

Suppose now that α is a literal. The case that $\mathcal{D}$ is of length 1 was already covered.

Suppose now the statement holds for all previously derived sequents of $\mathcal{D}$. Consider the last inference of $\mathcal{D}$. E.g.,

$$\frac{\Sigma \mid \Gamma, B, C, \alpha \mid\!\!\frac{\ \ }{\Theta}\ \Delta}{\Sigma \mid \Gamma, B \wedge C, \alpha \mid\!\!\frac{\ \ }{\Theta}\ \Delta} \wedge\vdash$$

By the inductive hypothesis there is a proof of $\Sigma \mid \Gamma, B, C \mid\!\!\frac{\ \ }{\Theta}\ \Delta, \beta$. We extend this proof by an application of $\wedge\vdash$ to $\Sigma \mid \Gamma, B \wedge C \mid\!\!\frac{\ \ }{\Theta}\ \Delta, \beta$.

In general we use the following scheme for single-premise rules. If the last rule of $\mathcal{D}$ introduces α on the left side via weakening:

$$\frac{\begin{matrix}\mathcal{D}' \\ \vdots \\ \Sigma \mid \Gamma \mid\!\!\frac{\ \ }{\Theta}\ \Delta\end{matrix}}{\Sigma \mid \Gamma, \alpha \mid\!\!\frac{\ \ }{\Theta}\ \Delta} \mathsf{LW}$$

we use the proof

$$\frac{\begin{matrix}\mathcal{D}' \\ \vdots \\ \Sigma \mid \Gamma \mid\!\!\frac{\ \ }{\Theta}\ \Delta\end{matrix}}{\Sigma \mid \Gamma \mid\!\!\frac{\ \ }{\Theta}\ \Delta, \beta} \mathsf{RW}$$

Note here that weakening is the only way to introduce a literal as a principal formula.

If the last inference of $\mathcal{D}$ is not weakening and where $\mathcal{D}$ is a one-premise inference of the form:

$$\frac{\begin{matrix}\mathcal{D}' \\ \vdots \\ \Sigma' \mid \Gamma', \alpha \mid\!\!\frac{\ \ }{\Theta}\ \Delta\end{matrix}}{\Sigma \mid \Gamma, \alpha \mid\!\!\frac{\ \ }{\Theta}\ \Delta} \mathrm{R}$$

we have:

$$\begin{array}{cl} \Sigma' \mid \Gamma', \alpha \models_{\Theta'} \Delta' & \\ \cdots\cdots\cdots\cdots\cdots\cdots & \models\neg \quad \text{ind. hyp.} \\ \Sigma' \mid \Gamma' \models_{\Theta'} \Delta', \beta & \\ \cdots\cdots\cdots\cdots\cdots\cdots & \mathrm{R} \\ \Sigma \mid \Gamma \models_{\Theta} \Delta, \beta & \end{array}$$

For 2-premise rules the argument is analogous and left to the reader.

Suppose now α is complex. The case in which $\mathcal{D}$ is of length 1 was already considered. We move to the inductive step and consider a proof $\mathcal{D}$ of length greater than 1. Consider first cases in which A is a side-formula of the last inference of $\mathcal{D}$. E.g.,

$$\frac{\Sigma \mid \Gamma, B, C, \alpha \mid\!\!\frac{\ \ }{\Theta}\ \Delta}{\Sigma \mid \Gamma, B \wedge C, \alpha \mid\!\!\frac{\ \ }{\Theta}\ \Delta} \wedge\vdash$$

By the inductive hypothesis there is a proof of $\Sigma \mid \Gamma, B, C \vdash_{\Theta} \Delta, \beta$. We extend this proof by an application of $\wedge\vdash$ to $\Sigma \mid \Gamma, B \wedge C \vdash_{\Theta} \Delta, \beta$. Other rules are handled similarly.

In general we use the following scheme. We consider now the case of a 2-premise rule R. Suppose $\mathcal{D}$ has the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ \Sigma \mid \Gamma_1, \alpha \vdash_{\Theta} \Delta_1\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots\\ \Sigma \mid \Gamma_2, \alpha \vdash_{\Theta} \Delta_2\end{array}}{\Sigma \mid \Gamma, \alpha \vdash_{\Theta} \Delta}\,\mathrm{R}$$

we have:

$$\dfrac{\dfrac{\Sigma \mid \Gamma_1, \alpha \models_{\Theta} \Delta_1}{\Sigma \mid \Gamma_1 \models_{\Theta} \Delta_1, \beta}\,\models\neg \quad \text{ind. hyp.} \qquad \dfrac{\Sigma \mid \Gamma_2, \alpha \models_{\Theta} \Delta_2}{\Sigma \mid \Gamma_2 \models_{\Theta} \Delta_2, \beta}\,\models\neg \quad \text{ind. hyp.}}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \beta}\,\mathrm{R}$$

The case of single-premise rules is handled in an analogous way to the inductive base.

Consider now cases in which A is principal in the last inference of $\mathcal{D}$.

Suppose $A = B \wedge C$. Then the last inference of $\mathcal{D}$ is of the form:

$$\dfrac{\Sigma \mid \Gamma, B, C \vdash_{\Theta} \Delta}{\Sigma \mid \Gamma, B \wedge C \vdash_{\Theta} \Delta}\,\wedge\vdash$$

By the inductive hypothesis (over the length of A) there are proofs of $\Sigma \mid \Gamma, B \vdash_{\Theta} \Delta, \neg C$ and $\Sigma \mid \Gamma \vdash_{\Theta} \Delta, \neg B, \neg C$. We apply to the latter proof $\vdash\neg\wedge$ to obtain $\Sigma \mid \Gamma \vdash_{\Theta} \Delta, \neg(B \wedge C)$.

In general we use the following scheme. We consider a two-premise rule Ra (where Rb is the corresponding β-rule). Suppose $\mathcal{D}$ is of the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ \Sigma \mid \Gamma, \alpha_1 \vdash_{\Theta} \Delta\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots\\ \Sigma \mid \Gamma, \alpha_2 \vdash_{\Theta} \Delta\end{array}}{\Sigma \mid \Gamma, \alpha \vdash_{\Theta} \Delta}\,\mathrm{Ra}$$

We have:

$$\dfrac{\dfrac{\Sigma \mid \Gamma, \alpha_1 \models_{\Theta} \Delta}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \beta_1}\,\models\neg \quad \text{ind. hyp.} \qquad \dfrac{\Sigma \mid \Gamma, \alpha_2 \models_{\Theta} \Delta}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \beta_2}\,\models\neg \quad \text{ind. hyp.}}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \beta}\,\mathrm{Rb}$$

The single-premise case is handled analogously. ■

Properties 3 and 4

We show, where $i \in \{1, 2\}$:

$$\frac{\Sigma \mid \Gamma \models_{\Theta} \Delta, \alpha}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \alpha_i} \quad \text{and} \quad \frac{\Sigma \mid \Gamma \models_{\Theta} \Delta, \beta}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \beta_1, \beta_2}$$

From this we immediately get, where $i, j \in \{1, 2\}$,

$$\cfrac{\cfrac{\Sigma \mid \Gamma \models_{\Theta} \Delta, \alpha}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \alpha_i}}{\Sigma \mid \Gamma \models_{\Theta} \Delta, \alpha_i, \beta_j}\ \mathsf{RW}$$

We show the property via an induction over the length of the proof of $\Sigma \mid \Gamma \vdash_{\Theta} \Delta, A$ where $A = \alpha$ respectively $A = \beta$.

Base case: Suppose $s = \Sigma \mid \Gamma \vdash_{\Theta} \Delta, A$ is an axiom. Note that in this case A is a literal and so the property holds trivially.

Inductive Step. Suppose now $\mathcal{D}$ is of length $n + 1$ and the last sequent derived is $s = \Sigma \mid \Gamma \vdash_{\Theta} \Delta, A$.

Suppose first that A is the principal formula in the last derivation. We consider the various available rules.

First assume $\mathcal{D}$ has the form:

$$\cfrac{\begin{array}{c}\mathcal{D}' \\ \vdots \\ \Sigma \mid \Gamma \vdash_{\Theta} \Delta\end{array}}{\Sigma \mid \Gamma \vdash_{\Theta} \Delta, A}\ \mathsf{RW}$$

Then consider the proof, where $A = \alpha$ and $i \in \{1, 2\}$,

$$\cfrac{\begin{array}{c}\mathcal{D}' \\ \vdots \\ \Sigma \mid \Gamma \vdash_{\Theta} \Delta\end{array}}{\Sigma \mid \Gamma \vdash_{\Theta} \Delta, \alpha_i}\ \mathsf{RW}$$

and, where $A = \beta$,

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}'\\ \vdots\\ \Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta\end{array}}{\Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, \beta_1}\,\mathsf{RW}}{\Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, \beta_1, \beta_2}\,\mathsf{RW}$$

Suppose now the last rule is a two-premise rule. We consider the case that $A = \alpha$. Then $\mathcal{D}$ has the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ \Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, \alpha_1\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots\\ \Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, \alpha_2\end{array}}{\Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, A}\,\mathrm{Ra}$$

In this case $\mathcal{D}_1$ and $\mathcal{D}_2$ are the proofs we are looking for. The other principal cases are similar and left to the reader.

Now consider cases where A is a side formula of the last derivation of $\mathcal{D}$. Considering one-premise rules, $\mathcal{D}$ has the form:

$$\dfrac{\Sigma' \mid \Gamma' \mid\!\!\frac{}{\Theta'}\ \Delta', A}{\Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, A}\,\mathrm{R}$$

By the inductive hypothesis, there is a proof $\mathcal{E}$ of $\Sigma' \mid \Gamma' \mid\!\!\frac{}{\Theta'}\ \Delta', \alpha_1$ and of $\Sigma' \mid \Gamma' \mid\!\!\frac{}{\Theta'}\ \Delta', \alpha_2$ in case $A = \alpha$, and of $\Sigma' \mid \Gamma' \mid\!\!\frac{}{\Theta'}\ \Delta', \beta_1, \beta_2$ in case $A = \beta$. In the latter case, consider now the proof:

$$\dfrac{\begin{array}{c}\mathcal{E}\\ \vdots\\ \Sigma' \mid \Gamma' \mid\!\!\frac{}{\Theta'}\ \Delta', \beta_1, \beta_2\end{array}}{\Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, \beta_1, \beta_2}\,\mathrm{R}$$

In the previous case consider the proof, where $i \in \{1, 2\}$,

$$\dfrac{\begin{array}{c}\mathcal{E}\\ \vdots\\ \Sigma' \mid \Gamma' \mid\!\!\frac{}{\Theta'}\ \Delta', \alpha_i\end{array}}{\Sigma \mid \Gamma \mid\!\!\frac{}{\Theta}\ \Delta, \alpha_i}\,\mathrm{R}$$

The case of two-premise rules is similar and left to the reader.

This completes our proof. ■

Property 2

In the following we show, where $\Sigma, \Sigma', \Gamma, \Gamma' \nvDash \Theta, \Theta'$,

$$\frac{\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{=}} \Delta, \alpha \quad \Sigma' \mid \Gamma' \mathrel{|\!\!\underset{\Theta'}{=}} \Delta', \beta}{\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{=}} \Delta, \Delta'} \vDash\top$$

Let in the following $\mathcal{D}$ be a *cut*-free proof of $s = \Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{-}} \Delta, \alpha$ and $\mathcal{E}$ be a *cut*-free proof of $t = \Sigma' \mid \Gamma' \mathrel{|\!\!\underset{\Theta'}{-}} \Delta', \beta$.

We show the property via an induction on the complexity of α (respectively of β).

Suppose first α is a literal. Without loss of generality suppose α is an atom while $\beta = \neg\alpha$. We now show that the property holds for the literal α by an induction on the length of $\mathcal{D}$ and $\mathcal{E}$.

In the base case both $\mathcal{D}$ and $\mathcal{E}$ are introductions of axioms.

In case both s and t are proper axioms $\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{-}} \Delta, \Delta'$ is also a proper axiom due to the closure of $\mathbb{P}$ under cut and contraposition.

Suppose now that both are logical axioms. Note that in this case $\Sigma = \emptyset = \Theta$. Consider first the case of axioms of the form $s = C \mathrel{|\!\!\underset{\emptyset}{-}} C$ and $t = \neg C \mathrel{|\!\!\underset{\emptyset}{-}} \neg C$ then $\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{-}} \Delta, \Delta' = C, \neg C \mathrel{|\!\!\underset{\emptyset}{-}} C, \neg C$. We obtain it as follows:

$$\dfrac{\dfrac{C, \neg C \mathrel{|\!\!\underset{\emptyset}{-}}}{C, \neg C \mathrel{|\!\!\underset{\emptyset}{-}} C}\,\mathsf{LW}}{C, \neg C \mathrel{|\!\!\underset{\emptyset}{-}} C, \neg C}\,\mathsf{RW}$$

Other case are similar and left to the reader.

We now consider the mixed case in which s is a proper axiom and t is a logical axiom. In case t is of the form $\beta \mathrel{|\!\!\underset{\emptyset}{-}} \beta$ the claim follows since by contraposition also $\Sigma \mid \Gamma, \beta \mathrel{|\!\!\underset{\Theta}{-}} \Delta$ is a proper axiom. Suppose $t = \ \mathrel{|\!\!\underset{\emptyset}{-}} \alpha, \beta$, then $\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\underset{\Theta,\Theta'}{-}} \Delta, \Delta' = \Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{-}} \Delta, \alpha = s$. Thus the claim holds trivially. The other mixed case is analogous.

We now move to proofs $\mathcal{D}$ and $\mathcal{E}$ which are in sum greater than length 2. Note first that the only way to introduce a literal on the right side of a sequent is via weakening and axiom introduction. Without loss of generality suppose $\mathcal{D}$ is of length greater than 1. If α is the principal formula of the last inference of $\mathcal{D}$ then the last step is weakening and $\mathcal{D}$ is of the form

$$\dfrac{\begin{array}{c}\mathcal{D}' \\ \vdots \\ \Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{-}} \Delta\end{array}}{\Sigma \mid \Gamma \mathrel{|\!\!\underset{\Theta}{-}} \Delta, \alpha}\,\mathsf{RW}$$

Then the proof we are seeking is as follows (where “Weakening” covers applications of all the available weakening rules DE, BE, RW, LW):

$$\dfrac{\begin{array}{c}\mathcal{D}'\\ \vdots \\ \Sigma \mid \Gamma \mathrel{|\!\!\!\frac{}{\Theta}} \Delta\end{array}}{\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\!\frac{}{\Theta, \Theta'}} \Delta, \Delta'}\ \text{Weakening}$$

Now suppose α is not principal in the last inference of $\mathcal{D}$. Suppose the last inference is a two-premise inference and so $\mathcal{D}$ has the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots \\ \Sigma \mid \Gamma_1 \mathrel{|\!\!\!\frac{}{\Theta}} \Delta_1, \alpha\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots \\ \Sigma \mid \Gamma_2 \mathrel{|\!\!\!\frac{}{\Theta}} \Delta_2, \alpha\end{array}}{\Sigma \mid \Gamma \mathrel{|\!\!\!\frac{}{\Theta}} \Delta, \alpha}\ \text{R}$$

We have:

$$\dfrac{\dfrac{\Sigma \mid \Gamma_1 \mathrel{|\!\!\!\frac{}{\Theta}} \Delta_1, \alpha \quad \Sigma' \mid \Gamma' \mathrel{|\!\!\!\frac{}{\Theta'}} \Delta', \beta}{\Sigma, \Sigma' \mid \Gamma_1, \Gamma' \mathrel{|\!\!\!\frac{}{\Theta, \Theta'}} \Delta_1, \Delta'}\ \models\top \quad \text{IH} \qquad \dfrac{\Sigma \mid \Gamma_2 \mathrel{|\!\!\!\frac{}{\Theta}} \Delta_2, \alpha \quad \Sigma' \mid \Gamma' \mathrel{|\!\!\!\frac{}{\Theta'}} \Delta', \beta}{\Sigma, \Sigma' \mid \Gamma_2, \Gamma' \mathrel{|\!\!\!\frac{}{\Theta, \Theta'}} \Delta_2, \Delta'}\ \models\top \quad \text{IH}}{\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\!\frac{}{\Theta, \Theta'}} \Delta, \Delta'}\ \text{R}$$

Other cases are similar and left to the reader.

We now move to the case in which α is a non-literal.

By properties 3 and 4, there are proofs of $\Sigma \mid \Gamma \mathrel{|\!\!\!\frac{}{\Theta}} \Delta, \alpha_1, \beta_2$, of $\Sigma \mid \Gamma \mathrel{|\!\!\!\frac{}{\Theta}} \Delta, \alpha_2$, and of $\Sigma' \mid \Gamma' \mathrel{|\!\!\!\frac{}{\Theta'}} \Delta', \beta_1, \beta_2$. By the inductive hypothesis on the length of α there is a proof of $\Sigma, \Sigma' \mid \Gamma' \mathrel{|\!\!\!\frac{}{\Theta'}} \Delta, \Delta', \beta_2$ given $\Sigma \mid \Gamma \mathrel{|\!\!\!\frac{}{\Theta}} \Delta, \alpha_1, \beta_2$, and $\Sigma' \mid \Gamma' \mathrel{|\!\!\!\frac{}{\Theta'}} \Delta', \beta_1, \beta_2$. Also by the inductive hypothesis, there is a proof of $\Sigma, \Sigma' \mid \Gamma, \Gamma \mathrel{|\!\!\!\frac{}{\Theta, \Theta'}} \Delta, \Delta'$ given $\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\!\frac{}{\Theta, \Theta'}} \Delta, \Delta', \beta_2$ and $\Sigma, \Sigma' \mid \Gamma, \Gamma' \mathrel{|\!\!\!\frac{}{\Theta, \Theta'}} \Delta, \Delta', \alpha_2$.

This completes our proof. ∎

Appendix B. Proof of $\mathbb{P}$-Relevance for $\mathbf{SC}_{\mathbb{P}}^{\Theta}$

Definition 23. $\mathsf{Sub}^{+}(\Gamma)$ is the closure of Γ under $\pi^{+} : \mathcal{L} \to \wp(\mathcal{L})$ where $A \mapsto \begin{cases}\{\alpha_1, \alpha_2\} & A = \alpha \\ \{\beta_1, \beta_2\} & A = \beta\end{cases}$ and $\mathsf{Sub}^{+}_{\mathsf{Lit}}(\Gamma) = \mathsf{Sub}^{+}(\Gamma) \cap \mathsf{Lit}$.

Example 12. Where $\Gamma = \{p \wedge \neg(s \vee q), \neg(r \wedge u)\}$ then $\mathsf{Sub}^{+}_{\mathsf{Lit}}(\Gamma) = \{p, \neg s, \neg q, \neg r, \neg u\}$.

Fact 5. $\mathsf{Val}(\Gamma) = \mathsf{Sub}^{+}_{\mathsf{Lit}}(\Gamma)$.

Fact 6. If $\Gamma \mid_{\mathbb{P}} \Gamma'$ then $\Theta \mid_{\mathbb{P}} \Theta'$ for any $\Theta \in \wp(\mathsf{Sub}^{+}(\Gamma))$ and $\Theta' \in \wp(\mathsf{Sub}^{+}(\Gamma'))$.

Proof. Note that $\Gamma \mid_{\mathbb{P}} \Gamma'$ iff for all $l \in \mathsf{Val}(\Gamma)$ and for all $l' \in \mathsf{Val}(\Gamma')$, (1) there is no $\langle \Sigma, X, \Theta, Y \rangle \in \mathbb{P}$ for which $l \in X$ and $l' \in Y$ and (2) $\mathsf{Atoms}(l) \cap \mathsf{Atoms}(l') = \emptyset$. From this we immediately get $\mathsf{Val}(\Gamma) \mid_{\mathbb{P}} \mathsf{Val}(\Gamma')$ and, where $\Theta \subseteq \mathsf{Sub}^{+}(\Gamma)$ and $\Theta' \subseteq \mathsf{Sub}^{+}(\Gamma')$, since $\mathsf{Val}(\Theta) \subseteq \mathsf{Val}(\Gamma)$ and $\mathsf{Val}(\Theta') \subseteq \mathsf{Val}(\Gamma')$ also $\Theta \mid_{\mathbb{P}} \Theta'$. ∎

Proposition 9 follows as a corollary of Proposition 25 below.

Proposition 25 (Subformula Property)**.** Where $\mathcal{D}$ is a *cut*-free derivation of $\Sigma \mid \Gamma \mathrel{|\!\overline{\,{}_{\Theta}\,}} \Delta$ and

(1) A occurs in Γ' for some sequent $\Sigma' \mid \Gamma' \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'$ in $\mathcal{D}$, then $A \in \mathsf{Sub}^+(\Gamma)$.
(2) A occurs in Δ' for some sequent $\Sigma' \mid \Gamma' \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'$ in $\mathcal{D}$, then $A \in \mathsf{Sub}^+(\Delta)$.

Proof. We show inductively that for every sub-derivation $\mathcal{D}'$ of $\mathcal{D}$ with conclusion $\Sigma' \mid \Gamma' \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'$ and every sub-conclusion $\Sigma'' \mid \Gamma'' \mathrel{|\!\overline{\,{}_{\Theta''}\,}} \Delta''$ of $\mathcal{D}'$ we have $\Gamma'' \subseteq \mathsf{Sub}^+(\Gamma')$ and $\Delta'' \subseteq \mathsf{Sub}^+(\Delta')$.

For sub-derivations of length 1, namely axiom introductions, this is trivial.

For the inductive step we distinguish cases according to the last inference rule R in $\mathcal{D}'$. We give several examples and leave other cases to the reader.

Suppose $\mathsf{R} = \mathsf{LW}$. Then $\mathcal{D}'$ is of the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ \Sigma' \mid \Gamma'' \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'\end{array}}{\Sigma' \mid \Gamma'', A \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'}\,\mathsf{LW}$$

Let $\Sigma''' \mid \Gamma''' \mathrel{|\!\overline{\,{}_{\Theta'''}\,}} \Delta'''$ be a sub-conclusion of $\mathcal{D}_1$. Then, by the inductive hypothesis, $\Gamma''' \subseteq \mathsf{Sub}^+(\Gamma'')$ and $\Delta''' \subseteq \mathsf{Sub}^+(\Delta')$. Clearly, then also $\Gamma''' \subseteq \mathsf{Sub}^+(\Gamma'' \cup \{A\})$.

Suppose now $\mathsf{R} = \wedge\vdash$. Then $\mathcal{D}'$ is of the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ \Sigma' \mid \Gamma_1, A, B \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'\end{array}}{\Sigma' \mid \Gamma_1, A \wedge B \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'}\,\wedge\vdash$$

Let $\Sigma''' \mid \Gamma''' \mathrel{|\!\overline{\,{}_{\Theta'''}\,}} \Delta'''$ be a sub-conclusion of $\mathcal{D}_1$. By the inductive hypothesis, $\Gamma''' \subseteq \mathsf{Sub}^+(\Gamma_1 \cup \{A, B\})$ and $\Delta''' \subseteq \mathsf{Sub}^+(\Delta')$. Since $A, B \in \mathsf{Sub}^+((\Gamma_1 \cup \{A \wedge B\})$, also $\Gamma''' \subseteq \mathsf{Sub}^+((\Gamma_1 \cup \{A \wedge B\})$.

Suppose now $\mathsf{R} = \neg\wedge\vdash$. Then $\mathcal{D}'$ is of the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ \Sigma_1 \mid \Gamma_1, \neg A \mathrel{|\!\overline{\,{}_{\Theta_1}\,}} \Delta_1\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots\\ \Sigma_2 \mid \Gamma_2, \neg B \mathrel{|\!\overline{\,{}_{\Theta_2}\,}} \Delta_2\end{array}}{\Sigma_1, \Sigma_2 \mid \Gamma_1, \Gamma_2, \neg(A \wedge B) \mathrel{|\!\overline{\,{}_{\Theta_1,\Theta_2}\,}} \Delta_1, \Delta_2}\,\neg\wedge\vdash$$

Let $\Sigma''' \mid \Gamma''' \mathrel{|\!\overline{\,{}_{\Theta'''}\,}} \Delta'''$ be a sub-conclusion of $\mathcal{D}_1$ or of $\mathcal{D}_2$. By the inductive hypothesis, $\Gamma''' \subseteq \mathsf{Sub}^+(\Gamma_1 \cup \{\neg A\}) \cup \mathsf{Sub}^+(\Gamma_2 \cup \{\neg B\})$ and $\Delta''' \subseteq \mathsf{Sub}^+(\Delta_1) \cup \mathsf{Sub}^+(\Delta_2)$. Since $\neg A, \neg B \in \mathsf{Sub}^+(\Gamma_1 \cup \Gamma_2 \cup \{\neg(A \wedge B)\})$, also $\Gamma''' \subseteq \mathsf{Sub}^+(\Gamma_1 \cup \Gamma_2 \cup \{\neg(A \wedge B)\})$.

Other cases are similar and left to the reader. ■

Lemma 26. *Where $\mathcal{D}$ is a cut-free derivation of $\Sigma \mid \Gamma, \Gamma' \mathrel{|\!\overline{\,{}_{\Theta}\,}} \Delta$ and $\Gamma \cup \Delta \mid_{\mathbb{P}} \Gamma'$, each sub-conclusion in $\mathcal{D}$ is of the form $\Sigma' \mid \Omega, \Omega' \mathrel{|\!\overline{\,{}_{\Theta'}\,}} \Delta'$ where $\Omega \subseteq \mathsf{Sub}^+(\Gamma)$, $\Omega' \subseteq \mathsf{Sub}^+(\Gamma')$, $\Delta' \subseteq \mathsf{Sub}^+(\Delta)$, and $\Omega' \mid_{\mathbb{P}} \Gamma \cup \Delta$.*

Proof. Let $\Sigma' \mid \Psi \vdash_{\Theta'} \Delta'$ be a sub-conclusion of $\mathcal{D}$. By Proposition 25, $\Delta' \subseteq \mathsf{Sub}^+(\Delta)$ and $\Psi \subseteq \mathsf{Sub}^+(\Gamma) \cup \mathsf{Sub}^+(\Gamma')$. Note that since $\Gamma' \not\Vdash_{\mathbb{P}} \Gamma \cup \Delta$, by Fact 6, also $\mathsf{Sub}^+(\Gamma) \cup \mathsf{Sub}^+(\Delta) \not\Vdash_{\mathbb{P}} \mathsf{Sub}^+(\Gamma')$ and thus $(\mathsf{Sub}^+(\Gamma) \cup \mathsf{Sub}^+(\Delta)) \cap \mathsf{Sub}^+(\Gamma') = \emptyset$. The rest follows immediately. ■

Proposition 27 (Basic $\mathbb{P}$-Relevance of $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$)**.** If $\Theta \neq \emptyset$, $\Sigma \mid \Gamma, \Gamma' \models_{\Theta} \Delta$ and $\Gamma \cup \Delta \not\Vdash_{\mathbb{P}} \Gamma'$ also $\Sigma \mid \Gamma \models_{\Theta} \Delta$.

Proof. Suppose $\Sigma \mid \Gamma, \Gamma' \vdash_{\Theta} \Delta$ is provable in $\mathsf{SC}^{\boldsymbol{\Theta}}_{\mathbb{P}}$. By Lemma 5, there is a *cut*-free proof $\mathcal{D}$ of it. The proof strategy in the following is to define a transformation π of $\mathcal{D}$ into a proof of $\Sigma \mid \Gamma \vdash_{\Theta} \Delta$ in an inductive manner.

Note that by Lemma 26,

($\star$) each derived sequent in a subproof $\mathcal{D}'$ of $\mathcal{D}$ is of the form $\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'}, \Gamma'_{\mathcal{D}'} \vdash_{\Theta_{\mathcal{D}'}} \Delta_{\mathcal{D}'}$ where $\Gamma'_{\mathcal{D}'} \not\Vdash_{\mathbb{P}} \Gamma \cup \Delta$ and $\Gamma_{\mathcal{D}'} \subseteq \mathsf{Sub}^+(\Gamma)$, $\Gamma'_{\mathcal{D}'} \subseteq \mathsf{Sub}^+(\Gamma')$, and $\Delta_{\mathcal{D}'} \subseteq \mathsf{Sub}^+(\Delta)$.

Consider first sub-proofs $\mathcal{D}'$ of $\mathcal{D}$ that are axiom introductions. Suppose $\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'} \vdash_{\Theta_{\mathcal{D}'}} \Delta_{\mathcal{D}'}$ is a proper axiom. By ($\star$), $\Delta_{\mathcal{D}'} \subseteq \mathsf{Sub}^+(\Delta)$ and $\Gamma_{\mathcal{D}'} \subseteq \mathsf{Sub}^+(\Gamma) \cup \mathsf{Sub}^+(\Gamma')$. Also, each $l \in \Gamma_{\mathcal{D}'}$ is relevant to each $l' \in \Delta_{\mathcal{D}'}$ and thus to some $l' \in \mathsf{Sub}^+(\Delta)$. Since $\Gamma' \not\Vdash_{\mathbb{P}} \Gamma \cup \Delta$ and hence, by Fact 6, $\mathsf{Sub}^+(\Gamma') \not\Vdash_{\mathbb{P}} \mathsf{Sub}^+(\Gamma) \cup \mathsf{Sub}^+(\Delta)$, $l \in \mathsf{Sub}^+(\Gamma)$. This shows that $\Gamma_{\mathcal{D}'} \subseteq \mathsf{Sub}^+(\Gamma)$. So we let $\pi(\mathcal{D}') = \mathcal{D}'$.

Suppose now $\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'} \vdash_{\Theta_{\mathcal{D}'}} \Delta_{\mathcal{D}'}$ is a logical axiom. Then $\Sigma_{\mathcal{D}'} = \Theta_{\mathcal{D}'} = \emptyset$. Note that the axiom is either of the form $C \vdash_{\emptyset} C$ or $\vdash_{\emptyset} C, \neg C$. (Since $\Sigma \mid \Gamma, \Gamma' \models_{\Theta} \Delta$ is undefeated and $\mathcal{D}$ is *cut*-free, the axiom cannot be of the form $C, \neg C \vdash_{\emptyset}$.) Suppose the former. By ($\star$), $C \in \mathsf{Sub}^+(\Delta) \cap (\mathsf{Sub}^+(\Gamma) \cup \mathsf{Sub}^+(\Gamma'))$ and since $\Gamma' \not\Vdash_{\mathbb{P}} \Gamma \cup \Delta$, $\mathsf{Sub}^+(\Delta) \cap \mathsf{Sub}^+(\Gamma') = \emptyset$ and hence $C \in \mathsf{Sub}^+(\Gamma) \cap \mathsf{Sub}^+(\Delta)$. We let $\pi(\mathcal{D}') = \mathcal{D}'$. In the latter case, by ($\star$), $C, \neg C \in \mathsf{Sub}^+(\Delta)$ and we let $\pi(\mathcal{D}') = \mathcal{D}'$.

For the inductive step consider the last inference rule of $\mathcal{D}'$ is a single premise rule and so $\mathcal{D}'$ has the form:

$$\dfrac{\begin{array}{c}\mathcal{D}''\\ \vdots \\ \Sigma_{\mathcal{D}''} \mid \Gamma_{\mathcal{D}''}, \Gamma'_{\mathcal{D}''} \vdash_{\Theta_{\mathcal{D}''}} \Delta_{\mathcal{D}''}\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'}, \Gamma'_{\mathcal{D}'} \vdash_{\Theta_{\mathcal{D}'}} \Delta_{\mathcal{D}'}}\ \mathsf{R}$$

By the inductive hypothesis, $\pi(\mathcal{D}'')$ yields a proof of $\Sigma_{\mathcal{D}''} \mid \Gamma_{\mathcal{D}''} \vdash_{\Theta_{\mathcal{D}''}} \Delta_{\mathcal{D}''}$. We distinguish now different cases according to the nature of the rule R to determine $\pi(\mathcal{D}')$.

Where $\mathsf{R} = \mathsf{BE}$ and $\Sigma_{\mathcal{D}'} = \Sigma_{\mathcal{D}''} \cup \{A\}$, let $\pi(\mathcal{D}')$ be

$$\dfrac{\begin{array}{c}\pi(\mathcal{D}'')\\ \vdots \\ \Sigma_{\mathcal{D}''} \mid \Gamma_{\mathcal{D}''} \vdash_{\Theta_{\mathcal{D}''}} \Delta_{\mathcal{D}''}\end{array}}{\Sigma_{\mathcal{D}''}, A \mid \Gamma_{\mathcal{D}''} \vdash_{\Theta_{\mathcal{D}''}} \Delta_{\mathcal{D}''}}\ \mathsf{BE}$$

The case for $\mathsf{R} = \mathsf{DE}$ is similar.

Suppose R is a rule where the principal formula is in $\Gamma_{\mathcal{D}'}$ or in $\Delta_{\mathcal{D}'}$. Then let $\pi(\mathcal{D}')$

be:

$$\dfrac{\begin{array}{c}\pi(\mathcal{D}'')\\ \vdots \\ \Sigma_{\mathcal{D}''} \mid \Gamma_{\mathcal{D}''} \mid\!\underline{}_{\Theta_{\mathcal{D}''}}\ \Delta_{\mathcal{D}''}\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'} \mid\!\underline{}_{\Theta_{\mathcal{D}'}}\ \Delta_{\mathcal{D}'}}\,\mathsf{R}$$

Suppose R is a rule where the principal formula is in $\Gamma'_{\mathcal{D}'}$. Then let $\pi(\mathcal{D}') = \pi(\mathcal{D}'')$. We proceed with two-premise rules.

Suppose $\mathcal{D}'$ is of the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots \\ \Sigma_{\mathcal{D}_1} \mid \Gamma_{\mathcal{D}_1}, \Gamma'_{\mathcal{D}_1} \mid\!\underline{}_{\Theta_{\mathcal{D}_1}}\ \Delta_{\mathcal{D}_1}, D_1\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots \\ \Sigma_{\mathcal{D}_2} \mid \Gamma_{\mathcal{D}_2}, \Gamma'_{\mathcal{D}_2} \mid\!\underline{}_{\Theta_{\mathcal{D}_2}}\ \Delta_{\mathcal{D}_2}, D_2\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'}, \Gamma'_{\mathcal{D}'} \mid\!\underline{}_{\Theta_{\mathcal{D}'}}\ \Delta_{\mathcal{D}'}, D}\,\mathsf{R}$$

Then $\pi(\mathcal{D}')$ is:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots \\ \Sigma_{\mathcal{D}_1} \mid \Gamma_{\mathcal{D}_1} \mid\!\underline{}_{\Theta_{\mathcal{D}_1}}\ \Delta_{\mathcal{D}_1}, D_1\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots \\ \Sigma_{\mathcal{D}_2} \mid \Gamma_{\mathcal{D}_2} \mid\!\underline{}_{\Theta_{\mathcal{D}_2}}\ \Delta_{\mathcal{D}_2}, D_2\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'} \mid\!\underline{}_{\Theta_{\mathcal{D}'}}\ \Delta_{\mathcal{D}'}, D}\,\mathsf{R}$$

Suppose now $\mathcal{D}'$ is of the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots \\ \Sigma_{\mathcal{D}_1} \mid \Gamma_{\mathcal{D}_1} \setminus \{D_1\}, D_1, \Gamma'_{\mathcal{D}_1} \mid\!\underline{}_{\Theta_{\mathcal{D}_1}}\ \Delta_{\mathcal{D}_1}\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots \\ \Sigma_{\mathcal{D}_2} \mid \Gamma_{\mathcal{D}_2} \setminus \{D_2\}, D_2, \Gamma'_{\mathcal{D}_2} \mid\!\underline{}_{\Theta_{\mathcal{D}_2}}\ \Delta_{\mathcal{D}_2}\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'} \setminus \{D\}, D, \Gamma'_{\mathcal{D}'} \mid\!\underline{}_{\Theta_{\mathcal{D}'}}\ \Delta_{\mathcal{D}'}}\,\mathsf{R}$$

Then $\pi(\mathcal{D}')$ is:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots \\ \Sigma_{\mathcal{D}_1} \mid \Gamma_{\mathcal{D}_1} \setminus \{D_1\}, D_1 \mid\!\underline{}_{\Theta_{\mathcal{D}_1}}\ \Delta_{\mathcal{D}_1}\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots \\ \Sigma_{\mathcal{D}_2} \mid \Gamma_{\mathcal{D}_2} \setminus \{D_2\}, D_2 \mid\!\underline{}_{\Theta_{\mathcal{D}_2}}\ \Delta_{\mathcal{D}_2}\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'} \setminus \{D\}, D \mid\!\underline{}_{\Theta_{\mathcal{D}'}}\ \Delta_{\mathcal{D}'}}\,\mathsf{R}$$

Suppose $\mathcal{D}'$ is of the form:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots \\ \Sigma_{\mathcal{D}_1} \mid \Gamma_{\mathcal{D}_1}, \Gamma'_{\mathcal{D}_1} \setminus \{D_1\}, D_1 \mid\!\underline{}_{\Theta_{\mathcal{D}_1}}\ \Delta_{\mathcal{D}_1}\end{array} \quad \begin{array}{c}\mathcal{D}_2\\ \vdots \\ \Sigma_{\mathcal{D}_2} \mid \Gamma_{\mathcal{D}_2}, \Gamma'_{\mathcal{D}_2} \setminus \{D_2\}, D_2 \mid\!\underline{}_{\Theta_{\mathcal{D}_2}}\ \Delta_{\mathcal{D}_2}\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'}, \Gamma'_{\mathcal{D}'} \setminus \{D\}, D \mid\!\underline{}_{\Theta_{\mathcal{D}'}}\ \Delta_{\mathcal{D}'}}\,\mathsf{R}$$

Then $\Sigma_{\mathcal{D}'} = \Sigma_{\mathcal{D}_1} \cup \Sigma_{\mathcal{D}_2}$, $\Gamma_{\mathcal{D}'} = \Gamma_{\mathcal{D}_1} \cup \Gamma_{\mathcal{D}_2}$, $\Delta_{\mathcal{D}'} = \Delta_{\mathcal{D}_1} \cup \Delta_{\mathcal{D}_2}$ and $\pi(\mathcal{D}')$ is, where "Weakening" refers to multiple applications of BE, DE, RW, and LW:

$$\dfrac{\begin{array}{c}\pi(\mathcal{D}_1)\\ \vdots \\ \Sigma_{\mathcal{D}_1} \mid \Gamma_{\mathcal{D}_1} \mid\!\!\frac{\quad}{\Theta_{\mathcal{D}_1}} \Delta_{\mathcal{D}_1}\end{array}}{\Sigma_{\mathcal{D}'} \mid \Gamma_{\mathcal{D}'} \mid\!\!\frac{\quad}{\Theta_{\mathcal{D}'}} \Delta_{\mathcal{D}'}}\text{Weakening}$$

Other cases are similar and left to the reader. ■