Proof-Theoretic Semantics and the Interpretation of Atomic Sentences

PRESTON STOVALL[1]

Abstract: This essay addresses one of the open questions of proof-theoretic semantics: how to understand the semantic values of atomic sentences. I embed a revised version of the explanatory proof system of Millson and Straßer (2019) into the proof-theoretic semantics of Francez (2015) and show how to specify (part of) the intended interpretations of atomic sentences on the basis of their occurrences in the premises and conclusions of inferences to and from best explanations.

Keywords: Proof-Theoretic Semantics, Atomic Sentences, Intended Interpretation, Explanation

1 Introduction

Proof-theoretic semantics (PTS) is an approach toward meaning that uses rules for inferring to and from linguistic or logical expressions as a basis for semantic evaluation. To date, however, there has been little discussion of the semantic values of—or the rules that govern—atomic sentences in PTS. Their values are generally either assumed given from outside the proof system (e.g., Francez, Dyckhoff, & Ben-Avi, 2019, and Francez, 2015), or provided by the inferences from their own assumptions as in (Francez, 2017), or stipulated via definitional systems of the sort found in (Prawitz, 1973). Francez (2015, p. 377) considers this one of the open questions within proof-theoretic semantics: how should we understand the proof-theoretic semantic values of the atoms of an interpreted language?

[1]This essay forms a triad with (Stovall, 2019) and (Stovall, in press) which collectively represent an attempt to rewrite the first two chapters of my dissertation. I am indebted to many people for taking the time to talk with me and offer comments on this material. I am particularly grateful for feedback from Robert Brandom, Ulf Hlobil, Daniel Kaplan, Jared Millson, Nissim Francez, Jaroslav Peregrin, Ivo Pezlar, Mark Risjord, and Shawn Standefer at various times over the last few years. I would also like to thank two reviewers for this journal, and the participants at the 2019 Logica conference. Work on this article was supported by the joint Lead-Agency research grant between the Austrian Science Foundation (FWF) and the Czech Science Foundation (GAČR), *Inferentialism and Collective Intentionality*, GF17-33808L.

In this essay I provide an answer to that question by showing how to use a revised version of the explanatory proof system of Millson and Straßer (2019) to individuate introduction and elimination rules for atomic sentences, under an intended interpretation, in terms of their roles in explanation under that interpretation. Within the context of (Francez, 2015), this determines a proof-theoretic semantic value for the atoms of an interpreted language, just as a specific assignment of truth conditions to sentences fixes atomic sentence meaning in model-theoretic semantics. While this doubtlessly does not exhaust everything one could mean by 'proof-theoretic meaning', it provides a formally tractable mechanism for thinking about at least one dimension of linguistic meaning: explanatory role.

2 Overview of the proof-theoretic semantics of Francez

2.1 Conventions

Consider a language $\mathscr{L}$ consisting of a countable set of atoms and the recursively defined set of sentences derived from the atoms and the Boolean operators in the usual way. Let uppercase Latin letters (except X and Y) range over sentences of $\mathscr{L}$, lowercase Latin letters range over atoms, uppercase Greek letters range over finite sets of sentences, and X and Y range over finite sets of literals (atoms and their negations). I abuse notation and use metalinguistic variables as illustrating instances in the text of the essay.

For display of logical relations I use a natural deduction system of the sort employed in (Francez, 2015). Rules of introduction and elimination are associated with each logical operator. These are *rules of inference* in that they determine which inferences can be made to and from the logically complex sentences of $\mathscr{L}$: the introduction rules for an operator * specify the conditions under which a sentence having * as a major operator can be inferred to as the conclusion of an inference, and the elimination rules for * specify the conditions under which such a sentence can be inferred from as a premise in an inference. I treat inferences as single-step derivations in this essay, rather than as acts of inferring, and I understand a derivation is an ordered series of applications of these rules of inference displayed as a tree structure. Thus, each application of a rule will count as an *inference* from the premise(s) to the conclusion and a *derivation* of the conclusion from the premises.

The object-language expressions that occur in the premises and conclusions of applications of rules of inference in these derivation systems are sequents of the form $\Gamma : A$, where the antecedent Γ is a set of sentences that encodes assumptions on which the derivation of the succedent A may depend. This presentation is in what Francez calls *logistic* form, and it has the virtue of displaying the background contexts that the application of a rule may depend on (Francez gives Γ more structure, but that is not relevant here). The rule $\rightarrow I$, for instance, reads as follows:

$$\frac{\Gamma, A : B}{\Gamma : A \rightarrow B} \rightarrow I$$

As a natural deduction system (rather than a sequent calculus) elimination rules operate on succedents rather than antecedents, as can be seen with the elimination rules for the conditional and conjunction:

$$\frac{\Gamma : A \qquad \Gamma : A \rightarrow B}{\Gamma : B} \rightarrow E \qquad \frac{\Gamma : A \wedge B}{\Gamma : A} \wedge EL \qquad \frac{\Gamma : A \wedge B}{\Gamma : B} \wedge ER$$

This notation allows for using the intuitive rules of natural deduction while preserving the expressive power of sequent calculi in the ability to keep track of the context from which some derivable formula is derivable. This will in turn facilitate tracking various features of context that attend explanatory inferences. Francez (2015) examines a variety of classical and non-classical natural deduction systems; nothing I say here turns on adopting any particular one.

2.2 Overview of PTS

In PTS the semantic values of sentences are delimited by their canonical derivations. For logically complex sentences, these are determined by the introduction and elimination rules for the logical operators (cf. Francez, 2015, Def. 1.5.9; the major premise of an inference rule for a logical operator '*' is the premise containing '*').

Definition 1 (Canonical Derivations of Logically Complex Sentences From Open Assumptions) *Let A be a logically complex sentence.*

1. *A derivation for Γ : A is I-canonical for A iff it satisfies one of the two following conditions:*
   - *The last rule applied in the derivation is an I-rule for the main operator of A.*

- *The last rule applied in the derivation is an assumption-discharging E-rule, the major premise of which is some B in* Γ*, and its encompassed sub-derivations are all canonical derivations of A.*

2. *A derivation for* $\Gamma, A : B$ *is E-canonical for A iff it starts with an application of an E-rule with A as the major premise.*[2]

The second condition for I-canonical derivations ensures that the commutativity and associativity of disjunction count as part of the meaning of disjunctions. The proof-theoretic semantic value for a sentence *A* at a context Γ can now be defined as follows (cf. Francez, 2015, p. 38):

Definition 2 (I-Canonical and E-Canonical Comprehension)

- *The I-canonical comprehension of A at* $\Gamma =_{def}$ *the collection of all I-canonical derivations for A from* Γ.

- *The E-canonical comprehension of A at* $\Gamma =_{def}$ *the collection of all E-canonical derivations for A from* Γ, A.

The use of the term 'comprehension' is meant to signal a contrast with the extensional semantics of model theory. Notice that PTS is hyperintensional: the compositionality of derivations ensures that the I-canonical derivations for *A* will differ from those for $A \wedge A$ (cf. the discussion in Pezlar, 2018).

2.3 PTS and atomic sentences

The definition for comprehension is general with regard to atoms and logically complex sentences. The notion of canonical derivability, however, applies only to logically complex sentences. For atoms we have (recall that lowercase Latin letters range over the atoms of $\mathscr{L}$):

Definition 3 (Canonical Derivations of Atomic Sentences From Open Assumptions)

1. *A derivation for* $\Gamma : p$ *is I-canonical for p iff the last rule applied in the derivation is an I-rule for p.*

[2]It is possible to loosen this definition and allow that a derivation is E-Canonical for a formula *A* just in case the first rule applied to *A* as a major premise is an E-rule for *A*. This would allow that derivations containing inferences prior to the application of an E-rule for *A* as a major premise would still count as part of *A*'s meaning. My thanks to Nissim Francez for pointing this out.

2. *A derivation for* $\Gamma, p : A$ *is E-canonical for* p *iff it starts with an application of an E-rule with* p *as the major premise.*[3]

Once we have rules for introducing and eliminating atoms, this will determine their I-canonical and E-canonical derivations, which in turn will fix their comprehensions.

3 Explanation as a basis for introducing and eliminating atomic sentences

The notion of an *explanation* has some intuitive appeal for fixing (part of) the proof-theoretic semantic values of atomic sentences. Just as there are both introduction and elimination rules, so are there two orders or directions of explanation. On one hand we may keep a sentence fixed and look to see which contexts or circumstances better explain it—e.g., we might ask what explains the existence of Socrates. On the other hand we may consider sentences across different contexts and look to see which other sentences are better explained by it—and so we might ask what the existence of Socrates explains. Whatever else it is to understand what a sentence means, one who can specify what it explains and what explains it will know something of its meaning. One might wish to call this the 'explanatory comprehension' of atoms in PTS, as to distinguish the range of semantic content fixed by these explanations from other sorts of proof-theoretic meaning, but I suppress that here.[4]

Philosophical logicians are beginning to investigate proof-theoretic formalizations of explanatory inference (see Litland, 2017, Millson, Khalifa, & Risjord, 2018, Millson & Straßer, 2019, Poggiolesi, 2016 and Poggiolesi, 2018). Each of Litland (2017), Millson and Straßer (2019), and Poggiolesi (2016) use a proof-theoretic metalanguage containing explanatory inferences in order to provide introduction and elimination rules for object-language talk whose meaning has been historically difficult to render in precise terms: viz., factive and nonfactive ground (Litland), best explanation (Millson and Straßer), and formal explanation (Poggiolesi). Each proof system distinguishes two sorts of rules and the derivations they define: one set of rules

[3] A similar remark holds here as in footnote 2.

[4] One also might adopt an epistemological stance and consider the reasons we have for believing what we do, as opposed to the reasons there are for things being that way, but I will continue to speak in an ontological register.

governing inferences that are explanatory, and the other governing non-explanatory inferences. The latter set of rules delimits the fragment of the language defined by the logical operators. I will follow this approach here. Litland (2017, p. 283) refers to these as explanatory and plain arguments, Poggiolesi (2016, p. 3149), following Aristotle and Bolzano, as proofs-why and proofs-that, and Millson and Straßer (2019, pp. 128 and 135) consider explanatory arguments and a version of Smullyan's (1968) classical logic for the sequent calculus LK.

What is needed is an account that allows us to infer to and from some sort of explanation (grounding, best, formal, etc). Schematically this would be a pair of rules that tell us 1) that when one is entitled to infer A from some context Γ, and certain other conditions are met, then one can infer that Γ is the relevant sort of explanation for A, and 2) that when A is the relevant sort of explanation for something, and certain side conditions are met, we can infer A. That is, where $:^{\blacktriangleright}$ indicates that the antecedent explains (in whatever sense) the succedent, we want to fill in the side conditions in the following:

$$\frac{\Gamma : A \qquad side\,conditions}{\Gamma :^{\blacktriangleright} A} \qquad \frac{side\,conditions \qquad \Gamma, A :^{\blacktriangleright} B}{\Gamma : A}$$

To do so is to specify when one can infer *to* an explanation, and what one can infer *from* an explanation. The system of Millson and Straßer (2019) has the virtue of specifying side conditions for sorting candidate explanations in order to arrive at the best (if any).

4 The explanatory proof theory of Millson and Straßer

In this part of the essay I present a streamlined version of part 4 of (Millson & Straßer, 2019), though I have translated their sequent notation into a logistic natural deduction notation of the sort Francez uses. Millson and Straßer are interested in non-monotonic or defeasible explanations, and they begin with the idea of a defeasible inference, displayed as follows.

$$\Sigma \mid \Gamma :_{\Theta} A$$

There are two key technical devices in play here. First, inferences are evaluated relative to both an antecedent Γ whose content is explanatory and a background Σ of sentences against which an explanation is assessed but which are not playing a role in the explanation. This allows them to distinguish background information, which may not be relevant to an explanation, from that which is doing the explanatory work (consider inferring 'the match is lit' from 'the match is struck' in situations where 'today is Tuesday' is part of the background but is not relevant to the inference). Second, a defeater set Θ keeps track of whether an inference is defeated. Θ is a set of sets of sentences, and the inference

$$\Sigma \mid \Gamma :_{\Theta} A$$

is defeated anytime $\Sigma \cup \Gamma$ logically derives all of the members of any element of Θ.

Material axioms are given with antecedents and succedents as finite sets of literals and literals, respectively. Millson and Straßer integrate these material axioms with rules for the classical logical operators, and an inference $\Sigma \mid \Gamma :_{\Theta} A$ is logical, and hence indefeasible, just in case Θ is empty (see Millson & Straßer, 2019, Lemma 4.4). And because any antecedent Γ with logically complex sentences will, against a given background Σ, derive a set of literals, the material consequences of $\Gamma \cup \Sigma$ can be identified with the material consequences of that set. Some of these defeasible inferences will be explanations, and Millson and Straßer specify a method for sorting explanations so as to find the best. On this basis they give an introduction rule for inferring to a best explanation, and thereafter derive an elimination rule. I will sometimes refer to the antecedent of an explanation as the explanans and to the succedent as the explanandum.

Millson and Straßer take the notion of *sturdiness* as the key desideratum for a theory of best explanation. Roughly, an explanation is sturdy when its explanans would remain a good explanation for the explandum even if all of its competitor explanatia were false—understood as adding the negations of the sentences occurring in those competing explanatia to the background against which the explanation is assessed. There may be no such explanation (no inference from an explanans remains good upon supposition of the falsity of its alternatives). There may also be ties (multiple explanations remain undefeated when all of the competing explanatia are false). We use $\Xi_{\langle \Sigma, \Gamma, A \rangle}$ to denote the set of explanations that compete with Γ for best explaining A at Σ (see Definition 9 for the explicit definition of $\Xi_{\langle \Sigma, \Gamma, A \rangle}$). I will generally speak of a single best explanation in what follows, but it is important to remember that ties are possible. As we will see, sturdiness amounts to something like an introduction rule for best explanations.

Unfortunately, the use of $\Xi_{\langle \Sigma, \Gamma, A \rangle}$ as a means of testing for sturdiness admits a class of counterexamples (see Part 5 for discussion). To avoid this problem I will eventually revise their account of sturdiness, but I begin by laying out that account.

4.1 Inferring to a best explanation

Our task is to determine whether an explanation is the best by determining whether it is sturdy. Where $\Sigma \,|\, \Gamma :_{\Theta} A$ is the explanation in question, consider the class of potential competitor explanations for A at Σ. This is the set of candidate best explanations for A at Σ, denoted $\mathbf{S}_{\langle \Sigma, A \rangle}$ and referred to as an Antecedent Set (cf. Millson & Straßer, 2019, Definition 4.8).

Definition 4 (Antecedent Set, $\mathbf{S}_{\langle \Sigma, A \rangle}$) *For any* $\varnothing \subset \Sigma, \Theta, A \subset \mathscr{L}$, *let*

$$\boldsymbol{S}_{\langle \Sigma, A \rangle} =_{def} \{X \text{ such that } \varnothing \subset X \subset Lit \text{ and } \Sigma \,|\, X :_{\Theta \cup A} A \text{ is provable } \}.$$

Notice that the inclusion of A in the defeater set ensures that we consider only material inferences. To have an example to refer to throughout the discussion, suppose that we are looking for the best explanation for the existence of Socrates (A) at some context (Σ). The antecedent set $\mathbf{S}_{\langle \Sigma, A \rangle}$ will include as possible explanations, we may suppose, that his parents met, that they fell in love, that they were married, that a particular zygote (understood *de re*) was formed, etc. Depending on what information is contained in the background Σ, any one of these might be explanatory, though clearly some are better than others.

We arrive at competitor sets in two steps, first by using a principle to restrict $\mathbf{S}_{\langle \Sigma, A \rangle}$ to those that are the simplest, and then by using two principles to compare this restricted class of competitor explanations to Γ. Both steps are preserved in the revised notion of sturdiness introduced below, but the two principles employed at the second step are put to slightly different uses.

At the first step we invoke a principle that rules out those explanations that are more committive than necessary. We do this by removing any sets from $\mathbf{S}_{\langle \Sigma, A \rangle}$ that are supersets of other sets in $\mathbf{S}_{\langle \Sigma, A \rangle}$. The resulting set is denoted $\mathbf{S}^{\downarrow}_{\langle \Sigma, A \rangle}$. More generally (cf. Millson & Straßer, 2019, Definition 4.9; and recall that X and Y range over finite subsets of the literals of $\mathscr{L}$):

Definition 5 (Literal-Set Minimisation, $\mathbf{S}^{\downarrow}$) *For any* $\mathbf{S} \subset \mathscr{P}(Lit)$, *let*

$$\boldsymbol{S}^{\downarrow} =_{def} \boldsymbol{S} \backslash \{X \text{ such that } Y \in \boldsymbol{S} \text{ and } Y \subset X\}.$$

Continuing with our example of Socrates, suppose that $\mathbf{S}_{\langle \Sigma, A \rangle}$ includes both {a particular zygote was formed} and {a particular zygote was formed, {Socrates} exists}. In this case $\mathbf{S}^{\downarrow}_{\langle \Sigma, A \rangle}$ will include only the former explanans, on the principle that if it was explanatory at this background then additional information is otiose.

If the first step involves sorting good from bad explanations within $\mathbf{S}_{\langle\Sigma,A\rangle}$ according to a principle of simplicity, and arriving at the literal-set minimisation $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ as a result, the second step invokes two principles for comparing Γ with the elements of $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ so as to arrive at the right class of competing explanans.

The first principle Millson and Straßer enforce at the second step is another principle of simplicity. Intuitively, if an explanation X for A implies something Y that itself explains A, but which does not in turn imply X, then Y is the simpler explanation and so is preferable to X. And so just as literal set-minimalization removed supersets of other sets in $\mathbf{S}_{\langle\Sigma,A\rangle}$ in order to compare only the simplest explanations with Γ, we want to ensure that any explanation X in $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ that is simpler than Γ will undercut Γ as an explanans by defeating the inference from Γ to A. This defeat is ensured by the fact that, when we add the negations of every sentence in X to Σ, the resulting context-cum-explanans is incoherent. But if an explanation in $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ is more complicated than Γ, we want to be sure *not* to add its negations to the background when testing Γ so as not to defeat the inference from Γ by default.

For instance, where Γ is {a particular zygote was formed, Heraclitus exists} and $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ includes {a particular zygote was formed}, we want to add the negation of 'a particular zygote was formed' to Σ and ensure that the explanans Γ is undercut. By contrast, where Γ is {a particular zygote was formed} and $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ includes {a particular zygote was formed, Heraclitus exists}, we *do not* want to add the negations of the sentences in that latter set to Σ.

To enforce this notion of comparative simplicity we first define 'Δ is logically weaker than Γ' as follows (cf. Millson & Straßer, 2019, Definition 4.10; as a definition for *logical* weakness, notice that the absence of a defeater set indicates use of a logical consequence relation):

Definition 6 (Logically Weaker Than, $\Gamma >^{:} \Delta$) *For any* $\Gamma, \Delta \subset \mathscr{L}$, *let*

$$\Gamma >^{:} \Delta \textit{ iff } [\Gamma : \bigwedge\Delta \textit{ but not } (\Delta : \bigwedge\Gamma)].$$

We then require $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ (the defeater set) to include the following:

$$\bigcup\{X \in \mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle} \textit{ such that } \Gamma >^{:} X\}$$

That is, we require that the competitor set include any explantia X that are logically weaker than the candidate explanans Γ, meaning that Γ logically derives every formula in X but not vice versa. Thus when X is a subset of $\Xi_{\langle\Sigma,\Gamma,A\rangle}$, the negations of the sentences of $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ will include negations of formulae that are logically derivable from Γ. It follows that $\Sigma \cup \Gamma$ is incoherent, and so the inference to A will be defeated.

The second principle for comparing Γ with the minimal competitor explanations in $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ is a principle for evaluating competitor explanantia for Γ that are *not* logically weaker than Γ and so which should not straightaway defeat the inference to A. For some explanations Y in $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ are either logically stronger than Γ, in the sense that they imply every sentence in Γ without the converse implication holding, or they are neither logically weaker nor logically stronger than Γ. Of these explanations Y that are not logically weaker than Γ, we must ensure that when we consider whether Γ remains a good explanation on the supposition of their falsity that the explanation to Γ is not undercut simply in virtue of the fact that they share some literals in common with the set of literals implied by Γ. And so we must ensure that $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ does not contain any literals that are implied by Γ and which only occur in explanations Y that are not logically weaker than Γ. We do this by taking the union of the set of explanations Y in $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ that are not logically weaker than Γ and removing any literals that are logically derived by Γ.

Returning to the example concerning Socrates, where $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ includes $X = \{$a particular zygote was formed, Heraclitus exists$\}$, and supposing our candidate explanans is now {a particular zygote was formed, Socrates' parents fell in love}, the fact that X is not logically weaker than Γ means that we must remove the literal that X shares in common with Γ ('a particular zygote was formed') and add only the negation of 'Heraclitus exists' to Σ before seeing whether the explanation still holds. And this is the right result: when I consider whether Γ remains a good explanation even when the competing explanans X doesn't hold, I consider whether the formation of the zygote and his parents falling in love would still explain Socrates' existence if Heraclitus hadn't existed, as this is the only information that is new to that competing explanans (the inclusion of the sentence about his parents falling in love is to ensure that X is not already excluded according to Definition 6).

Let the literal consequence-set of Γ, denoted $\mathrm{Cn}_{Lit}(\Gamma)$ be defined as follows (cf. Millson & Straßer, 2019, Definition 4.11; once again note the restriction to the logical fragment of the consequence relation):

Definition 7 (Literal Consequence-Set, $\mathrm{Cn}_{Lit}(\Gamma)$) *For any* $\Gamma \subset \mathscr{L}$*, let*

$$Cn_{Lit}(\Gamma) =_{def} \{l \in Lit \text{ such that } \Gamma : l\}$$

And so we also include the following in $\Xi_{\langle\Sigma,\Gamma,A\rangle}$:

$$\Big(\bigcup\{Y \in \mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle} \textit{ such that } \Gamma \not>^{\because} Y\}\Big)\backslash Cn_{Lit}(\Gamma)$$

By barring from the competitor set $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ any literals that are consequences of Γ and are only had by explanations Y in $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ that are not logically weaker than Γ, we ensure that those competitors don't straightaway defeat the explanation from Γ when we add the negations of every sentence in $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ to the background and test Γ for sturdiness. Taking these three principles together we have the following general definition (cf. Millson & Straßer, 2019, Definition 4.12; this relation is not named there):

Definition 8 ($\mathbf{S} \setminus\!\setminus\ \Gamma$) *For any $X, Y \subset Lit$, $\Gamma \subset \mathscr{L}$ and $\mathbf{S} \subset \mathscr{P}(Lit)$, let*

$$\boldsymbol{S} \setminus\!\setminus\ \Gamma =_{def} \bigcup\{X \in \boldsymbol{S} \textit{ s.t. } \Gamma >^{\because} X\} \cup \Big(\Big(\bigcup\{Y \in \boldsymbol{S} \textit{ s.t. } \Gamma \not>^{\because} Y\}\Big)\backslash Cn_{Lit}(\Gamma)\Big)$$

This says that we arrive at $\mathbf{S} \setminus\!\setminus\ \Gamma$ by taking the union of 1) every set in $\mathbf{S}$ that is logically weaker than Γ, together with 2) the union of every set in $\mathbf{S}$ that is not logically weaker than Γ, but subtracting any literal that is a logical consequence of Γ. Consider the following (from Millson & Straßer, 2019, p. 143; more examples are given there):

$$\{\{\neg p\},\{q\},\{p,r\}\} \setminus\!\setminus\ \{p,q\} = \{q\} \cup \big(\{\neg p, p, r\}\backslash Cn_{Lit}(\{p,q\})\big) = \{q, \neg p, r\}$$

This yields the following definition for competitor sets (cf. Millson & Straßer, 2019, Definition 4.13):

Definition 9 (Competitor Set, $\Xi_{\langle\Sigma,\Gamma,A\rangle}$) *For any $\Sigma\,|\,\Gamma :_{\Theta} A$, let the competitor set of its antecedent Γ be*

$$\Xi_{\langle\Sigma,\Gamma,A\rangle} =_{def} \boldsymbol{S}^{\downarrow}_{\langle\Sigma,A\rangle} \setminus\!\setminus\ Cn_{Lit}(\Gamma)$$

This says that the competitor set for Γ is the set of formulae that are the result of taking the union of the minimized sets that individually materially (defeasibly) explain A against the background Σ, and subtracting any formulae that are literal consequences of Γ and which are implied only by elements of $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ that are not logically weaker than Γ.

To test whether Γ is the *best* explanation for A at Σ we then add the negation of all of the elements of this set (denoted $\neg\Xi_{\langle\Sigma,\Gamma,A\rangle}$) to the background and see whether the explanation still holds. If it does, one is entitled to infer that this is the best explanation. As noted above, this sturdiness rule amounts to something like an introduction rule for best explanations (this rule and the next are given at Millson & Straßer, 2019, p. 144):

$$\frac{\Sigma \,|\, \Gamma :_{\Theta,A} A \qquad \Sigma, \neg\Xi_{\langle \Sigma,\Gamma,A\rangle} \,|\, \Gamma :_{\Theta,A} A}{\Sigma \,|\, \Gamma :^{\blacktriangleright}_{\Theta,A} A} \; STR$$

Here the black triangle above the defeater set signifies that Γ is the best explanation for A at Σ.

If Sturdiness (*STR*) functions as a rule for inferring *to* a best explanation and is analogous to an introduction rule for best explanation, then Abduction (*ABD*) functions as a rule for inferring something *from* a best explanation and is analogous to an elimination rule for best explanation:

$$\frac{\Sigma \,|\, \Gamma, A :^{\blacktriangleright}_{\Theta} B \qquad \Sigma' \,|\, \Gamma' :_{\Psi,B} B}{\Sigma, \Sigma' \,|\, \Gamma, \Gamma' :_{\Theta,\Psi,B} A} \; ABD$$

5 A problem with Millson and Straßer's account, and a proposed solution

Only some best explanations are captured by Millson and Straßer's test for competitors, however. For anytime there are multiple minimal competitors that contain sentences that cannot all be false together, it will happen that the addition of the negations of the sentences in the competitor set to the background will lead to a contradiction. But from a contradiction anything follows, and this means that in such a case all of the members of some element of the defeater set will be derivable, thus defeating the inference. This can happen even when the competitor explanations are much poorer than the one they are competing against.

This can be seen with an example. Suppose we are evaluating the following explanations for the existence of Socrates: 1) a particular zygote was formed; 2) Socrates' parents fell in love and Heraclitus exists; 3) Socrates' parents fell in love and Heraclitus does not exist. It is clear that the first is the best explanation of this trio. But where

s = Socrates exists. z = a particular zygote was formed

h = Heraclitus exists. f = Socrates' parents fell in love

and using the familiar notation we have:

$\Gamma = \{z\}$ $\Sigma = \varnothing$

$\mathbf{S}^{\downarrow}_{\langle\Sigma,s\rangle} = \{\{z\}, \{f, h\}, \{f, \neg h\}\}$

$\Xi_{\langle\Sigma,\Gamma,s\rangle} = \left(\mathbf{S}^{\downarrow}_{\langle\Sigma,s\rangle} \setminus\!\setminus\ Cn_{Lit}(\Gamma)\right) = \left(\varnothing \cup \left(\{z, f, h, \neg h\}\setminus\{z\}\right)\right) = \{f, h, \neg h\}$

And now when we add the negations of the elements of $\Xi_{\langle\Sigma,\Gamma,s\rangle}$ (that is, f, h and $\neg h$) to Σ we have an incoherent set. As a result, the explanation from z to s is defeated.

The problem lies in the way the competitor set $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ is calculated. Millson and Straßer combine the elements of all of the sets in $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$ into one set (subtracting the literal consequences from those sets that are not logically weaker than Γ) and then test Γ against the negations of all of these sentences. Instead, we should test whether an explanation remains good on the supposition of the falsity of each of its competitors separately.[5] We do this as follows (with the superscripted R denoting that we have a revised notion of a competitor set).

Definition 10 (Revised Competitor Set, $\Xi^R_{\langle\Sigma,\Gamma,A\rangle}$) *For any $\Sigma \mid \Gamma :_\Theta A$, let the revised competitor set of its antecedent Γ be*

$$\Xi^R_{\langle\Sigma,\Gamma,A\rangle} =_{def} \left\{ X \text{ such that either } \begin{array}{l} X \in \boldsymbol{S}^{\downarrow}_{\langle\Sigma,A\rangle} \text{ and } \Gamma >^{:} X, \text{ or} \\ \text{there is some } Y \in \boldsymbol{S}^{\downarrow}_{\langle\Sigma,A\rangle} \text{ such that} \\ \Gamma \not>^{:} Y \text{ and } X = Y \setminus Cn_{Lit}(\Gamma) \end{array} \right\}$$

Whereas $\Xi_{\langle\Sigma,\Gamma,A\rangle}$ is a set of sentences, $\Xi^R_{\langle\Sigma,\Gamma,A\rangle}$ is a set of sets of sentences. This preserves the two-step procedure for arriving at the competitors against which an explanation is to be tested, but the two principles appealed to at the second step are now employed separately, on the elements of $\mathbf{S}^{\downarrow}_{\langle\Sigma,A\rangle}$, to define the revised competitor set. Where $\Xi^R_{\langle\Sigma,\Gamma,A\rangle} = X_1, \ldots, X_n$, and where $\neg X_i$

[5] A more complicated weighting could be given as well, of course; these are early days for proof-theoretic investigation into the conditions for inferring to and from best explanations.

denotes the negation of every sentence in X, we have the following revised notion of sturdiness (*RSTR*):

$$\frac{\Sigma \mid \Gamma :_{\Theta,A} A \qquad \Sigma, \neg X_1 \mid \Gamma :_{\Theta,A} A \qquad \ldots \qquad \Sigma, \neg X_n \mid \Gamma :_{\Theta,A} A}{\Sigma \mid \Gamma :^{\blacktriangleright}_{\Theta,A} A}\ RSTR$$

This definition delivers the right result with our example above, and using this basis for inferring *to* a best explanation, the rule for abduction does not need adjustment. I thus propose the following:

Definition 11 (Introduction and Elimination Rules for Atomic Sentences)

- *The introduction rules for an atom p are the applications of RSTR where p is the succedent in the conclusion of the application of the rule.*
- *The elimination rules for an atom p are the applications of ABD where p is in the succedent of the major premise of the application of the rule.*

According to Definition 3 in Section 2.3, this determines the set of canonical derivations for the atoms of $\mathscr{L}$, which thereby determines their semantic values according to Definition 2 in Section 2.2.

6 Summary

Proof-theoretic semantics offers a formally precise and philosophically illuminating investigation into areas of linguistic meaning and rule-governed rationality that have been dominated by model-theoretic semantics and representational notions of cognition over the last century. By establishing a beachhead into these areas on the basis of a semantics for atomic sentences that employs a framework for reasoning to and from best explanations, the possibility of productive proof-theoretic interventions into some of these debates is rendered more likely.

References

Francez, N. (2015). *Proof-theoretic semantics: Studies in logic, vol. 57*. London: College Publications.

Francez, N. (2017). On distinguishing proof-theoretic consequence from derivability. *Logique & Analyse*, *238*, 151–166.

Francez, N., Dyckhoff, R., & Ben-Avi, G. (2019). Proof-theoretic semantics for subsentential phrases. *Studia Logica*, *94*, 184–231.

Litland, J. (2017). Grounding ground. In K. Bennett & D. W. Zimmerman (Eds.), *Oxford Studies in Metaphysics, vol.10* (pp. 279–316). Oxford: Oxford University Press.

Millson, J., Khalifa, K., & Risjord, M. (2018). Inferentialist-expressivism for explanatory vocabulary. In V. K. O. Beran & L. Koreň (Eds.), *From rules to meanings: New essays in inferentialism* (pp. 155–178). New York: Routledge.

Millson, J., & Straßer, C. (2019). A logic for best explanations. *Journal of applied non-classical logics*, *29*(2), 124–128.

Pezlar, I. (2018). Proof-theoretic semantics and hyperintensionality. *Logique et analyse*, *61*(242), 151–161.

Poggiolesi, F. (2016). On defining the notion of complete and immediate formal grounding. *Synthese*, *193*(10), 3147–3167.

Poggiolesi, F. (2018). On constructing a logic for the notion of complete and immediate formal grounding. *Synthese*, *195*(3), 1231–1254.

Prawitz, D. (1973). Towards a foundation of a general proof theory. In A. J. P. Suppes L. Henkin & G. Moisil (Eds.), *Logic, methodology and philosophy of science IV* (Vol. 74, pp. 155–178). Amsterdam: North-Holland.

Smullyan, R. (1968). *First-order logic*. Berlin: Springer-Verlag.

Stovall, P. (2019). Characterizing generics are material inference tickets: A proof-theoretic analysis. *Inquiry*, published online March 18, 2019, 1–37.

Stovall, P. (in press). Essence as a modality: A proof-theoretic and nominalist analysis. *Philosophers' Imprint*.

Preston Stovall
University of Hradec Králové, Faculty of Philosophy and Social Science
The Czech Republic
E-mail: preston.stovall@uhk.cz