THE UNIVERSAL THEORY BUILDING TOOLKIT IS SUBSTRUCTURAL

SHAY ALLEN LOGAN

In some of his recent work, Jc Beall has argued for a position that, reduced to a slogan, amounts to the claim that logic is the universal theory building toolkit. In [8], I argued that Beall was mistaken about what logic this should lead us to accept. While my argument there is best understood as a reaction to Beall's proposal, the intellectual genealogy of both the technical and philosophical work I pursued there are naturally traced back to Alsadair Urquhart's 1972 paper 'Semantics for Relevant Logics' (SRL).[1] In fact, we can be more specific and trace it back to the following two-sentence fragment of SRL:

> 'In argumentation, we may have to consider not only what information may be available, but also what the facts are. That is, we conclude that a piece of information X determines p against a certain background of facts.'

In this brief passage, Urquhart drew attention to something very simple, but very important: argumentation does not occur in a vacuum; it occurs set against a background of facts.

But there's more! SRL not only provides the philosophical impetus for my work in [8], it also provides the core technical innovation I used there. *This* part can be traced back to a single sentence of SRL:

> 'It is clear that given any two pieces of information, X and Y, we may put them together to form a new piece of information, $X \cup Y$, containing all the information in X together with all the information in Y.'

The crucial idea introduced here is the idea that semantic notions might be interpreted using a binary operation on indices. This can be contrasted with, e.g. Kripke frames, where one uses a binary *relation* on indices instead.

Already in 1972, Urquhart had worked out much of the semantic picture this led to. In 1974, Kit Fine gave semantic theories very similar to Urquhart's that covered an even more diverse family of logics.[2] There is much to be said about both what stays the same and what changes in the move from Urquhart's semantics to Fine's. But for the sake of the story I'm telling, it suffices to highlight just one key interpretive difference: where Urquhart understood the indices in his models to play the role of *pieces of information*, Fine took the indices in his models to play the role of *theories*.

Appealing to theories rather than pieces of information gave rise to an immediate problem: if we understand indices to be *theories*, how should we interpret the binary operation we've defined on them? Urquhart, thinking of his indices as pieces of information, interpreted the binary operation he defined as set union. This isn't an option on Fine's theory-based understanding of the semantics since in general the union of two theories is *not* a theory. The alternative Fine proposed was to read the binary operation as the operation of closing one theory under another.

[1] See [14]. It's also worthwhile to mention that this is more of an autobiographical genealogy of this work, rather than a full tracing-back of the exact origins of the ideas. If we were pursuing the latter, we would probably have to go further back in time than we have here.

[2] See [6].

Fine's proposal looks nice from a distance, but on closer inspection can be seen to suffer from a rather alarming problem. If we write '$cl_{t_1}(-)$' (where we read '$-$' as 'blank') for the operation Fine reads as closing under t_1, then it will in general *not* be the case that $t_2 \subseteq cl_{t_1}(t_2)$ and it will in general *not* be the case that $cl_{t_1}(cl_{t_1}(t_2)) = cl_{t_1}(t_2)$. Thus what Fine was calling 'closure' was not, in fact, closure in any traditional sense. This left it unclear what the right philosophical interpretation of the binary operation at play in Fine's semantics should be—assuming there to even be one! The solution to this problem was floating in the air already by the time Fine's paper appeared in print. But the (to my mind at least) best presentation of it did not appear for another two decades.

In 'A General Logic,' John Slaney revisited the Urquhartian interpretation of logic as involving bodies of information.[3] Like Fine and Urquhart, Slaney equipped these sites of evaluation with a binary operation. Unlike either, he interpreted this operation as what he called the *application* of one theory to another.[4] Slaney characterized this operation by saying that when we apply the information t to the information u, we use the resources of t to say what inferences are available to us, and we use the resources of u to determine what is available for the inferences to act on. But this idea makes just as much sense in theory-land as it does in information-land: applying t to u roughly means taking t as one's *theory of entailment* and taking u as the theory to which the theory of entailment is applied. This *strongly* resonates with the Urquhartian idea to only examine inference 'as it occurs set against a background'. It also gives us a philosophically palatable way of understanding the binary operation at play in Fine's semantics.

The theory I proposed in [8] is a version of the theory these observations naturally lead us to. In particular, the *deep fried semantics* I presented there used a binary operation on a set of indices that I interpreted as theories. The binary operation was interpreted as *application* à la Slaney's proposal. After a few simplifying assumptions, the deep fried approach landed us at the well-known substructural relevant logic RW.

Thus deep fried semantics, which plays an important role in the discussion below, enjoys at least a deep spiritual connection to Alasdair Urquhart's work—albeit a connection that is mediated by Fine and Slaney along the way. But it's also true that deep fried semantics is a direct reaction against Beall's aforementioned arguments. In broad strokes, what Beall argues is that logic is that which is universal with regard to our theory building endeavors. Formally, Beall works this out by supposing we have on hand a range of consequence relations, and then seeing what they all agree on. The result, after some technical finagling, is that the universal theory building toolkit (and thus, for Beall, *logic*) is the weak, purely extensional relevant logic FDE.

I agree with Beall that we formally-minded folk ought to be examining what we might call the universal theory building toolkit. Unlike Beall, I will (in this paper at least) remain silent about whether we should take this theory to be logic. After all, regardless of whether the universal theory building toolkit is logic, it's clearly a philosophically important thing to be studying. But the question of whether one should take a stand on the issue of universal theory building toolkit being logic is a relatively minor thing to disagree about, so I take it that, philosophically speaking, Beall and I are largely on the same page.

Technically, however, the matter is different. In particular, I disagree with Beall about the universal theory building toolkit being FDE. As I pointed out a moment ago, following the Urquhart-Fine-Slaney route will land us at a substructural logic. Beall identifies FDE as

[3] See [13].

[4] As mentioned, this idea was in the air well before this piece, and Slaney acknowledges this in his paper. He explicitly cites [7], [9], [10], and Dunn's work on relevant implication in this regard.

the universal theory building toolkit, so is committed to saying the universal theory building toolkit is *not* substructural. This disagreement—the disagreement about structural rules—is the focus of the paper you're reading.

The source of this disagreement is a disagreement about the nature of the consequence relations. Beall says very little about them. I claim, and this is central to the work below, that consequence relations are themselves given to us by *theories of entailment*, and that these theories of entailment are theories that are as worthy of a spot at the table as any other theories might be. It's exactly in working our how to make room at the table for these theories that we find ourselves in need of the Urquhart-Fine-Slaney approach and, as a result, find ourselves embracing a substructural system. It also turns out, as I take time to remind readers at the end of this paper, that we have good, paradox-avoiding reasons to adopt such a system anyways.

A note on presentation: I've chosen to write the paper in the form of a dialogue. This is because it is essentially a compilation of dialogues I've had the pleasure to be a part of over the last few years of presenting this view—though, as it turns out, I respond much more nimbly in print than I do in person. Among the interlocutors worth personally thanking, Jc Beall deserves the most recognition, followed immediately by Eric Carter, Nathan Kellen, and Graham Leach-Krouse.

1. Some Troubling Inferences

Setting: It's early 2020, during the Covid-19 pandemic and global quarantine. An online meeting of the Logic Supergroup has just ended. One by one, faces disappear from the screen as people sign off. Soon only SL (a Substructural Logician) and TL (a Traditional Logician) remain on the screen. They're old friends, so a conversation naturally starts up.

TL: What an excellent talk! The speaker certainly had me convinced.
SL: I also enjoyed it, though I'm a bit less convinced. I detected a few troubling inferences in the middle of her talk that I'm not sure I can endorse.
TL: Really! I was paying close attention and didn't notice any such thing. What were the offending inferences?
SL: There were three inference forms that she seemed to be relying on that worried me. First, she seemed to allow herself to infer $\phi \to \psi$ from $\phi \to (\phi \to \psi)$. Second, she seemed to need to infer ψ from $(\phi \to \psi) \wedge \phi$. Finally, she seemed at several points to infer something of the form $(\phi_2 \to \phi_3) \to (\phi_1 \to \phi_3)$ from something of the form $\phi_1 \to \phi_2$. Or at least, this is how things seemed to me. I'm open to having misunderstood her.
TL: That's always a possibility, of course. But would it really be so bad if you were right? I don't see any issues at all with any of these inferences.
SL: Do you say that because you think she provided independent justification for these inferences being acceptable in the case at hand? Or is that you take these inferences to be generally acceptable in our theory building endeavors?
TL: The latter, I think, although I'm not sure what you mean by 'theory building endeavors'.
SL: I suppose I mean that you take these inferences to be part of the universal theory building toolkit.
TL: I'm still not sure I follow.
SL: Perhaps a toy example will help make things clear. Suppose we are studying creatures found in Eastern Australia. The members of one species we encounter – call it species X – have the following features:

(i) Members of species X have fur.
(ii) Members of species X are warm-blooded.
(iii) Members of species X lay eggs.

Suppose also that we're ill-equipped to the study we've set out to do and are, in fact, completely ignorant about Australia's fauna.

TL: Ill-equipped indeed!

SL: Right. But suppose we nonetheless set out to identify what phylogenetic class species X belongs to. If we do this using the information in an outdated biology textbook as our background theory, we may (in virtue of (iii)) come to the following conclusion:

(iv) Members of species X are not mammals.

If we use the information in a more up-to-date biology textbook as our background theory, we might instead come to the following conclusion:

(v) Members of species X *are* mammals – in particular, they are monotremes.

The lesson to learn here is this: what conclusions we can draw from a certain body of information we have on hand (e.g. from (i)-(iii)) will in general change when we change the background theory against which we draw said conclusions.[5]

TL: Ah, now I think I see where you're heading. You're pointing out that not *all* conclusions we can draw from a body of information exhibit this sort of background-dependence. For example, regardless of whether we use the outdated textbook or the up-to-date textbook, we will certainly be able to draw the following conclusions:

(vi) Members of species X have fur and members of species X are warm-blooded.
(vii) Members of species X have fur and members of species X lay eggs.

SL: Exactly this. It's worthwhile to state the contrast explicitly: (iv) and (v) followed from (i)-(iii) only in the presence of this or that background theory; (vi) and (vii) followed from (i)-(iii) simply in virtue of the fact that we were trying to build a *theory*, and not a mere *set of sentences*. Being a *theory* requires having a certain amount of regularity.

TL: Wait, hang on, I'm sorry. Shouldn't we be proceeding in the reverse order somehow? What I mean is that it seems odd to start out by thinking about theories and to work up from *there* to some sort of consequence relation.

SL: I don't think it's odd at all to proceed *from* theories *to* consequence relations and, indeed, that's just what I plan to do. The notion of a theory is, on the story I'm telling, the primitive notion in terms of which we will be defining other notions. This gives the story what I take to be a rather 'Brandom-ian' flavor, which I rather like.[6]

Regardless, recall that (as I mentioned earlier) I'm interested in what I called *the universal theory building toolkit*. This, as you might expect, is the collection of all inferences that are safe to use whenever we are building theories, no matter the subject matter of the theory. To put it another way, the universal theory building toolkit is the collection of inferences that all *actual* theories are invariant under.

TL: Invariant how?

SL: I imagine I'll have more to say about this later, so for the moment I'll just give a few hints. The basic idea is this: since theories exhibit certain amounts of regularity, it follows that there are inferences that, when applied to a theory, will keep you within that very theory,

[5]Gillman Payette has pointed out to me that there are traditions where 'information' is used in such a way that one cannot have information that is inaccurate. Were that the way 'information' was being used here, it would seem that the outdated biology textbook wasn't supplying information. So be it; I hereby declare that I am using 'information' in a way that does not carry this connotation.

[6]See [3] or [4].

not matter what the theory is. Put in more mathematical terms, we can imagine inferences as functions from (sometimes sets of) sentences to sentences. A theory is *invariant* under a given inference when, no matter which of its sentences we put into the inference, the sentence we get out was already in the theory.

TL: Ok. This is a familiar-enough way of thinking about invariance. It's also clear enough that the set containing all those inferences that literally all the theories are invariant under is (a) a reasonable thing to call by the title 'universal theory-building toolkit' and (b) an interesting thing worth studying.

All that said, I *still* don't see what's objectionable about the inferences in question. For example, no matter the context and no matter the subject matter of the theory we're building, it will *always* be acceptable to infer $\phi \to \psi$ from $\phi \to (\phi \to \psi)$, it seems to me. The other inferences seem similarly warranted. So I'm just not sure what's making you so grumpy.

2. Theory Building

SL: Let's first make sure we're on the same page when it comes to theory building. I recently read something that put it rather nicely; let's see if I can share my screen and show it to you.

A long, frustrating ordeal ensues. Some minutes later, a limited form of screen-sharing is achieved and ***SL*** *displays the following passage from* [2]*:*

> '[O]nce she has identified her target phenomenon (about which she aims to give the true and as-complete-as-possible theory), the task of the theorist is twofold:
> - gather the truths about the target phenomenon
> - construct the right closure relation to 'complete' the true theory – to give as full or complete a true theory as the phenomenon allows.'

TL: If I've understood the passage correctly, Beall thinks (and, I take it, you think) that theory building involves two intuitively different operations: the 'gather some truths' operation, and the 'close under a closure relation' operation. I'll admit I'd never taken the time to think hard about theory building, but it seems that Beall is correct to point out theory building does seem to depend on both of the operations he's identified.

SL: That's my reading as well. I also think that, when combined with a few uncontroversial theses, Beall's observation here is enough to give us good reason to claim that the inferences the speaker seemed to rely on can't be part of the universal theory building toolkit. So, unless she gave us some reason to think they're acceptable in the situation at hand (or I'm wrong about her relying on them), I think she's left her conclusion in some doubt.

TL: Hang on, I've lost track of the inferences you're troubled by. Would you kindly remind me what they were?

SL: Of course. Here, I'll write them on our shared screen.

Another long, frustrating ordeal ensues. Some minutes later, a limited form of screen-writing is achieved, though neither participant can write particularly elegantly on the shared platform. Still, ***SL*** *manages to write the following in mostly legible script:*

Rule 1: If $\phi \to (\phi \to \psi) \in th$, then $\phi \to \psi \in th$.
Rule 2: If $(\phi \to \psi) \wedge \phi \in th$, then $\psi \in th$.
Rule 3: If $\phi_1 \to \phi_2 \in th$, then $(\phi_2 \to \phi_3) \to (\phi_1 \to \phi_3) \in th$.

TL: I take it that '*th*' here is an arbitrary theory.

SL: That's right.

TL: And, just to be sure I'm on board, what you're saying is that none of these rules should be in the universal theory building toolkit. So you deny that the kind of regularity demanded by theory-hood would require us to add ψ to all theories that contain $(\phi \to \psi) \wedge \phi$
SL: That's correct.

***TL** makes an incredulous face.*

SL: I can see I have my work cut out for me when it comes to convincing you. I thought about it while we were settling our technical matters, and I think I only need the following four theses for my argument:

(1) Background theories are theories like any other.
(2) Within a context, the behavior of a given background theory is determined by features of the theory itself.
(3) The only type of feature relevant to determining how a background theory behaves in a context is the sentences it is committed to.
(4) The behavior of background theories under application is determined by a connective in the object language, rather than by something more mysterious.

TL: I see quite a bit I don't like here, my friend.
SL: Really? I'm surprised. Let's hear your objections to the theses first, then, before we turn to my argument, since it seems you're not likely to buy the argument if you don't buy these theses.

3. Application and Bunching

TL: While you were settling our technical problems and thinking about your theses, I took the time to skim the paper you mentioned above. The following passage is relevant to my first objection:

After a lag, the shared screen updates to show the following passage

> [T]heories are pictured as pairs (to highlight the closure relation):
>
> $$\langle T_1, \vdash_{T_1}\rangle, \langle T_2, \vdash_{T_2}\rangle, \langle T_3, \vdash_{T_3}\rangle, \ldots, \langle T_n, \vdash_{T_n}\rangle$$
>
> Logic shows up in each such theory-specific consequence relation $\vdash_{T_i}$; it is the relation under which all true theories, so understood, are closed; it is the relation on top of which all closure relations for our true theories are built.[7]

SL: Ah yes. I suppose I should have mentioned that what I've called 'the universal theory building toolkit' Beall has called 'the universal closure relation'.
TL: That's helpful, but it's not your *vocabulary* that's bothering me. What's bothering me is that your theses are all about background theories, and I see no mention of them here at all!
SL: Oh well that's also easy enough to clear up: what Beall calls 'closing under a closure relation' I call 'applying a background theory'. Does that make the connection clear enough?
TL: You've settled one of my worries only to raise another! Your first thesis says background theories are theories like any other. Now you're saying that background theories are closure relations. But this is clearly incorrect—a closure relation is a relation; a background theory is a theory. And (though I thought it went without saying) relations aren't theories and theories aren't relations.

[7] Beall's restriction to *true* theories strikes me as problematic, but in ways that are orthogonal to this paper. Thus, my interlocutors will put the issue aside.

SL: I see now I've been incautious in stating my position. Mea culpa. I don't want to say that background theories and consequence relations are the same thing; I want to say that the *operation* Beall has labeled 'closing under a closure relation' is an instance of the *operation* I call 'applying a background theory'.

TL: Ok good. At least you're not trying to convince me that relations are theories! That gives me some consolation. I'll admit, though, that it's not a lot, since it's not at all obvious to me that closing under a closure relation is an instance of whatever operation 'applying a background theory' is.

SL: It might help to think about what you do in your logic classroom. My guess is that you do something like the following: First, you tell your students that ϕ follows from Γ just if in no possibility is everything in Γ true while ϕ is untrue. Then you spell out in detail what the possibilities are and how to determine what's true in any one of them.

TL: That's exactly what I do. Given all this, students can then determine, for particular Γ and ϕ, whether ϕ follows from Γ—which is to say, whether ϕ is in the closure of Γ under the given consequence relation. No mention of theory application is required.

SL: On the contrary: there's an entirely clear case of theory application going on! When you teach your students what 'follows from' means and what the possibilities are and all that jazz, what you've done is teach your students a particular theory—namely, the classical theory of entailment. Closing a given set of sentences under the classical consequence relation is exactly the same thing as applying the classical theory of entailment, taken as a background theory, to determine what follows (classically) from the sentences at hand.

TL: Ah, I see. You take the theories of entailment themselves to be the theories that are being applied. I'll grant you this is *a* way of describing what goes on in my logic classroom. But this doesn't absolve you of wrongdoing, my friend. Theories of entailment will be theories that, perhaps among other things, contain sentences of the form 'this entails that'. These sentences are metalinguistic because 'entails' is not an object language expression. By saying that theories containing these sorts of expressions are theories like any others, you're mixing levels—that is, you're confusing metalanguage and object language.

SL: What can only be expressed in the metalanguage is determined by the choices we make about what to include in the object language. For example, it's not somehow determined in advance that conjunction has to be an object-language device. We can build languages where conjunction is missing. Were we discussing such a language 'and' would be a metalinguistic, rather an object-linguistic notion.

So what's relevant isn't what *is in fact* in the object language and in the metalanguage, but what we *should* include in the object language and in the metalanguage. Given that we're talking about the universal theory building toolkit, and that one of the theory building operations (application of one theory to another) *crucially* interacts with entailments, including entailment in the object language is a no-brainer.

TL: I don't know if I'd call it a no-brainer, but perhaps what you're saying is this: if we suppress the entailment connective, then it seems we have no recourse except to include it in the metalanguage. But since what follows from a given family of sentences does in fact depend on what background theory we are using, we will then need multiple entailment relations in our metalanguage. This leaves open a natural question: how do we in fact come to employ these different entailment relations? It seems there are two answers we might give (a) the entailment relations are just there—they're floating in the ether somehow, or perhaps we consult an oracle each time we put them to work—or (b) they're actually spelled out in some concrete way. Option (a) is, for obvious reasons, unappealing. But in case (b), we have to recognize that what's being spelled out is a *theory* governing the behavior

of the entailment relation at hand—in other words, what's being spelled out is a *theory of entailment*. And this, you're claiming, is a genuine theory that operates 'in the background'.
SL: That's a much better way to put it.
TL: Thanks! I'm not sure I'm totally convinced, but I'll at least grant that you've made a plausible case for the thesis that background theories are theories and that background theories and the background 'role' more generally are things we need to keep track of. Now I'm curious to hear what this thesis does for you—what role does it play in your argument?
SL: The most important thing to note is this: if background theories are theories like any other, it follows that background theories may themselves be constructed by applying a 'backbackground' theory to a 'forebackground' theory. Forebackground and backbackground theories, in turn, may have further internal structure, and this could go back arbitrarily (though presumably only arbitrarily *finitely*) far.
TL: Can you give an example of a backbackground theory? I'm having trouble picturing such a thing.
SL: Easy enough! On a toy way of understanding it, physics might be thought of as some axioms—say the axioms of quantum field theory or some such—together with their mathematical consequences. So physics, on this toy picture, is what we get when we apply our mathematical theory to a set of axioms.
TL: So far so good, but this is only two theories.
SL: Right. But we might also think of the mathematics we're applying here as itself being composed of some axioms (say, the ZFC axioms) together with all their logical consequences. So physics on this picture is what we get when we apply logic to mathematical axioms, then apply the resulting theory to our physical axioms.
TL: Oh I see. So in this example 'logic' shows up as a backbackground theory. I can see where forebackground theories and all the rest occur now, thanks.
SL: Excellent. So, like I was saying, what the first thesis forces us to realize is that theories can be constructed in very complex ways.
TL: I expect you have a candidate family of structures in mind that can track these different ways theories can be built.
SL: Right again. This time we'll steal a definition from Stephen Read.[8] Define the term 'bunch' together with the auxiliary terms 'I-bunch' and 'E-bunch' by simultaneous recursion as follows:

- Any sentence is a(n atomic) I-bunch.
- Any set of I-bunches is an E-bunch.
- I-bunches and E-bunches (and nothing else) are bunches.
- If X and Y are bunches, then $(X; Y)$ is an I-bunch.

TL: It'll take a minute for me to fully grok this definition. In the meantime, I have another question: how should I read the 'I-bunching' and 'E-bunching' operations?
SL: The idea is that bunches name theories. The atomic bunch 'ϕ' names the smallest theory containing ϕ. The theories named by more complex bunches are fairly straightforward:

- '$\{X_i\}_{i\in I}$' names the smallest theory that contains the theories named by each X_i;
- '$X; Y$' names the theory we get by applying the theory named by X to the theory named by Y.

TL: I think I get the picture. I only have one question left: suppose X and Y are bunches, and let $t_{X;Y}$ be the theory named by $X; Y$. Presumably, since $t_{X;Y}$ is a theory, it's just a set of sentences. And sets of sentences are perfectly good bunches. So $t_{X;Y}$ is also a bunch. But

[8] See [11].

then $X; Y$ and $t_{X;Y}$ are different bunches that name the same theory. Why do we want or allow this?
SL: This is an excellent question. My answer will, in some sense, just pass the buck: facts about how a theory *can be* built are matters of concern even in classical logic.
TL: Can you show me this 'concern' you speak of in action?
SL: Easily. Think, for example, of the usual conditional introduction rule:

$$\frac{X \cup \{A\} : B}{X : A \supset B}$$

TL: Wait, don't tell me! I think I see where you're going. Using your theory-goggles, you'll read 'X' as 'the theory containing all the formulas in X' and you'll read '$X \cup \{A\}$ as 'the theory containing A as well as all the formulas in X'. Thus, the fact that we *can build* a theory that (contains? entails? I'm not sure what word you'll use here) B by adding A to X is what tells us that the theory generated by X contains/entails/whatever $A \supset B$.
SL: Correct! And 'contains' will suffice. I'll say a bit more about that while dealing with the second thesis—speaking of which, are you ready to move on to the second thesis?
TL: Indeed.

4. Closure

SL: As a reminder, my second thesis is that, within a context, the behavior of a background theory is determined by features of the theory itself. The point, I should make clear, is that once we know what our background theory is and we know the context in which it's operating, we know everything we need to know about what it's going to do. We don't, in other words, have further variance about the turnstile, as Beall allows.
TL: That's what I took you to be saying, but I don't see how it can be correct! Are you honestly telling me that I can't vary the turnstile, as Beall does? Surely that's wrong, and wrong for reasons Beall is very clear about—in different theory building endeavors, we just *do* use different consequence relations.
SL: Well first, a note about what 'context' means. I don't think you've misrepresented what I mean, but I can easily see how what you're saying might be misunderstood, so I thought we should be clear.

A given theory t can occur as a foreground theory, as a background theory, as a back-background theory, etc. In general, given a bunch Γ, we can highlight (pick out, point at, whathaveyou) a given occurrence of the theory t within Γ. Supposing we write $\Gamma(t)$ for Γ with t highlighted in this way, then $\Gamma(-)$ (to be read 'Γ blank') is the context in which t occurs.
TL: Indeed, that's what I had in mind. I'm glad we're on the same page.
SL: Great! Then my answer to your question is that every instance of turnstile-variance examined by Beall is just an instance of context variance in this sense.
TL: Can you elaborate on that a bit?
SL: Of course. Recall that Beall writes '$\Gamma \vdash_{T_n} \phi$' to mean that ϕ follows from Γ according to the consequence relation at play in theory T_n. In brief, ϕ is a T_n-consequence of Γ.
TL: Correct. And he takes this to be synonymous with 'ϕ is in the T_n-closure of Γ.' Now, you've said that closing under a closure relation is the same thing as applying a theory of entailment. How does that play out here?
SL: It couldn't be easier! Let's let c_n be the theory of entailment underlying the closure relation $\vdash_{T_n}$. Then where Beall writes $\Gamma \vdash_{T_n} \phi$, I simply write $c_n; \Gamma \vdash \phi$. Here the 'naked' turnstile is simple containment—so $c_n; \Gamma \vdash \phi$ simply means that the theory we get by

applying c_n to Γ contains the sentence ϕ. Given everything that I've said so far, you should now be able to tell that $\Gamma \vdash_{T_n} \phi$ and $c_n; \Gamma \vdash \phi$ do in fact mean the same thing.

But despite having the same meaning, I think that the latter way of saying things is better. In particular, when we write $\Gamma \vdash_{T_n} \phi$, the theory c_n that we're using to define that consequence relation $\vdash_{T_n}$ is being squirreled away somewhat dishonestly in a subscript. When we write $c_n; \Gamma \vdash \phi$, in contrast, both the theory c_n and the role it is playing are being proudly acknowledged.

TL: I suppose being explicit about what we're doing is praiseworthy, sure. I'm not sure I'd go so far as to call Beall's use of subscripts 'dishonest squirreling away' of background theories, but that's a matter I'll let you take up with Beall. On that note, I'm wondering if you can make sense of what Beall *does* do? If so, I'd like to see it as it might help me make sense of what purpose your machinery is serving.

SL: I think I can. From my perspective, Beall seems to be interested in the question 'what follows from t no matter what background theory we use?' Stated otherwise, Beall examines the consequences of $b; t$ as we allow b to vary among all possible options. By cataloguing the set of inferences that are valid in this way, we get a theory that is universal in a certain minimal sense. The issue I have is that the framework I've provided makes his theory look odd in at least three ways.

TL: One of them I can already see: now that we have your bunching apparatus on hand, it's not clear why we should restrict our attention to those bunch-sentence inferences where the bunch in question has the form $b; t$. How else does it look odd to you?

SL: The first thing that bothers me concerns vocabulary. As a result of his suppression of background theories, Beall's account is left with no reason to take seriously connectives that directly interact with the application operation. I'll have more to say about this below, though, so I'll rush on and tell you the final thing that bothers me as well: not only does Beall restrict to bunches of the form $b; t$, he *also* requires that b be the type of theory that, when applied, results in a *closure relation*.

TL: Do you mean 'closure relation' in the Tarskian sense?

SL: Yes. The part I'm particularly grumpy about is that Beall has restricted the range of background theories he's examining to those that have the feature that for all t, we get both of the following:

- $b; t \supseteq t$;
- $b; (b; t) = b; t$.[9]

This is a further restriction on the sorts of theories Beall allows to play the 'b' role in '$b; t$'.

Altogether the point is that we should embrace a broader sort of universality than Beall has allowed. Our aim should be to describe what tools we can use when theory building *no matter the theory*. Our aim should not be to determine, as Beall did, what tools we can use to determine what follows from a given theory as we vary which closure-relation-generating background theory we apply to it.

5. Commitment and Containment

TL: I'll have to wait and see how the details pan out before I can evaluate all of that. I'm anxious to get to the good bits, though, so let's try to blast through these last two theses. First up is your claim that the only feature relevant to determining how a background theory behaves in a context is the sentences it is committed to.

[9]There is a mild abuse of language here—strictly speaking we ought to say that, e.g. the theory named by $b; t$ is contained in the theory named by t, etc.

SL: Looking at a toy example might help make clear what I have in mind here. So consider e.g. all the information we had about the motions of the planets as of 1900 or so. Now consider what follows from these when we, first, use classical mechanics as our background theory and, second, use relativistic mechanics as our background theory. It's well known, and a matter of import in the history of science, that the results will differ. We needn't worry about the exact nature of the difference. What's important for us is that it's not *inexplicable* why, even applied to the same data, these theories give different predictions. It's because the theories have different *commitments*.
TL: Thanks! That was quite helpful. The claim you seem to be making looks to have two components:

- Theories that behave differently do so because they actually differ in some way, and
- Given the kinds of things theories are, the only way they can differ is in what sentences they're committed to.

The first component I'm happy to endorse, but in order for me to endorse the second component, I'll first need to know more about the notion of *commitment* at play here.
SL: Easy enough—it's the same thing as containment. Theories just are sets of sentences.[10] I hope you'll agree that on this reading of commitment, your second component is entirely unobjectionable.
TL: Ah yes, you'd mentioned this already. I suppose that, given this reading, the second component *is* unobjectionable. But I'm a bit confused here—I thought we adopted bunches because theories *couldn't* be described as mere sets.
SL: Not quite. We adopted bunches because there were more *theory building operations* than the binary '$\in$' relation of set theory was capable of tracking (easily). At any rate, since theories are sets of sentences, it's hard to imagine what an objection to the third thesis would look like. It seems the only forceful way to object would be by presenting a counterexample—two theories that differ in behavior without differing in commitments. But if the theories don't differ in commitments, I wouldn't know how to tell them apart in the first place!
TL: Agreed. Let's move on to the final thesis, that the background behavior of a theory is determined by a connective in the object language, rather than by something more mysterious. We've already agreed that all behavior—and thus, in particular, the background behavior—of theories is determined by their content.What this thesis *seems* to be trying to do is point us in the direction of the exact sort of content we need to look at if we are interested in understanding how a theory is going to behave when we use it in the background. Apart from that, I'll admit I'm not completely sure about what it means.
SL: The point I'm trying to make with this thesis is that there is a certain *type* of sentence that we need to look to when we are determining how a theory behaves when used in the background.
TL: Can you say more? I'm still a bit at sea here.
SL: I suppose what I'm picturing is this: let t_1 and t_2 be theories. Suppose ψ is in the theory we get from applying t_1 to t_2. Presumably this happens because t_2 contains some bit of information ϕ that t_1 states is sufficient for ψ. This connection *seems* well-modeled by a connective that we write '$\phi \to \psi$'. All I'm claiming in my fourth thesis is that this idea—the idea that some '$\to$-like' connective mediates the way a theory behaves when we use it as a background theory—is correct. Given the way it behaves it's natural to call this connective the *entailment* connective and to call sentences dominated by this connective

[10] This is, I acknowledge, a very logician-y perspective to take on theories. There is, I think, room to take a more semantic view of theories—or at least, there will be room for such a thing once we have semantics in hand.

entailments. Thus, my fourth thesis can equally (but more succinctly) be put by saying this: if we want to know how a given theory is going to behave when used as a background theory, all we have to look at are its entailments.
TL: I don't see anything objectionable here, but again, I'd like to test my understanding. So, given what we've said so far, the application of a theory t_1 to a theory t_2 ought to be the theory we get from the following set of sentences:

$$\{\psi : \phi \to \psi \in t_1 \text{ and } \phi \in t_2\}$$

SL: That's exactly right! t_1's entailments ($\to$-formulas) determine how it behaves when used, as a background theory. And the way they determine its behavior is exactly as you've shown—they simply tell us what follows from what's in a given theory.
TL: I take it then that, on your point of view, any system that lacks a connective that behaves in this way is only telling part of the story when it comes to universal theory-building.
SL: That's correct! As I'm sure you recognize, this ties back to one of the criticisms I had of Beall's account from before: since the story Beall told completely ignored background theories, it naturally lacked any such connective.
TL: Got it. Now I'm interested to know what this fourth thesis does for you.
SL: The main thing it does is suggest what the semantic clause for the entailment connective ought to be.
TL: Wait, wait. Are you saying you can produce a full-fledged semantic theory based on these ideas? Talk about burying the lede! Let's see what this theory looks like; whatever objections I have left will certainly be easier to make once the actual theory is on the table.
SL: Very well. But first, a warning: what I'm going to give is *a* way of turning the ideas we've discussed so far into a semantics. The semantic clauses I'm about to show you are emphatically *not* here because I think they in fact capture the universal theory building toolkit in all its glory. The point is only to give a proof of concept; that is, to show that we can give a semantics that accords with the above intuitions, and that we could use such a thing to figure out what the universal theory building toolkit might be. If you think these are the wrong clauses, then you likely think they also lead us to the wrong universal theory building toolkit. That's fine. What matters is that you can see how to get from the basic picture involving theories and theory building to a semantics and from there to a universal theory building toolkit.
TL: Noted.

6. Semantics

SL: Ok great! So, first things first: let's say a **model** is a quadruple $\langle T, \ell, \sqsubseteq, \circ\rangle$ with

- T a set of of indices that we will think of as playing the role of *theories*;
- $\ell \in T$ is an index that we will think of as playing the role of the *universal* theory;
- $\sqsubseteq$ a partial ordering of T that we will think of as telling us when one theory *extends* another;
- $\circ$ a binary operation on T that we will think of as playing the role of application of one theory to another.

Given the roles we want $\sqsubseteq$ and $\circ$ to play, we should impose a few conditions on their interactions. In particular, if $a \sqsubseteq b$, then b should be a 'bigger' (or rather, shouldn't be a smaller) theory than a—that is, everything in a is already in b. So if we apply the same theory, say t to both a and b, then we should not end up with less in the b-case than we did

in the a-case. So whenever $a \sqsubseteq b$ is true, we expect $t \circ a \sqsubseteq t \circ b$ to also be true. A similar line of reasoning leads us to expect that whenever $a \sqsubseteq b$ is true, $a \circ t \sqsubseteq b \circ t$ will also be true.
TL: So far so good. But you'll also want to require $\circ$ to have the following features, right?

Commutativity: $a \circ b = b \circ a$
Associativity: $(a \circ b) \circ c = a \circ (b \circ c)$
Idempotence: $a \circ a = a$

SL: I want nothing to do with any of these! I'll admit that they all *look* friendly enough. But think about what they say and you'll see they're far from innocent.
TL: Alright, let's see. We wanted $\circ$ to play the role of application. So $a \circ b$ is, intuitively, the theory generated by taking a as background theory and applying it to b taken as foreground theory. Thus in $a \circ b$, it's a that is giving us our theory of entailment. In an abuse of Beall's notation, we might think of $a \circ b$ as the theory we get by closing b under the closure relation $\vdash_a$. And now I see that you're correct—there's in general no good reason to suppose that closing b under $\vdash_a$ gets us the same theory as we get by closing a under $\vdash_b$.
SL: Excellent reasoning! Of course, I'm worried that '$\vdash_a$' and '$\vdash_b$' (to the extent that I endorse such things at all) won't even be closure relations in the traditional, Tarskian sense. But that's a different point altogether. In any case, Associativity and Idempotence turn out to be equally suspicious when subjected to a similar inspection.
TL: I feel like once we work out *all* the details these problems will go away, but I can certainly see how they're at least *prima facie* problems, so I'll go along with things for the moment. I suppose the next thing you'll tell me is how to get from models to truth?
SL: Not quite! There's another matter we need to address first: let's say that a is *b-invariant* when $b \circ a = a$. The thing we're interested in is the *universal* theory-building toolkit—the tools we can use in all our theory building.

The reason I'm bringing this up is the following: we want ℓ to be a theory that *every* theory is invariant under. Thus ℓ will be universal among theories in the sense that applying ℓ always only gets us all and only the stuff we already have. So if we use the tools in ℓ in our theory building, we will for sure not be going astray—it will contain only those tools that all actual theories are invariant under. All told this suggests that we demand ℓ be a left-identity for $\circ$—that is, we require that $\ell \circ t = t$ for all t.
TL: I see! ℓ, then, is the index at which exactly those formulas that belong to the universal theory building toolkit turn out true, yes?
SL: Yes. And truth, as you probably expect, requires valuations.[11] To make things easy, we'll take valuations to record both which atomic sentences are true-in-a-theory ($\vDash_1$) and which atomic sentences are false-in-a-theory ($\vDash_0$).[12] Thus a *valuation* is a function from theories to functions from atoms to $\{\{1\}, \{0\}, \{0, 1\}, \emptyset\}$. Intuitively, valuations record which atoms a theory *says are true* and which atoms it *says are false*. The only condition we should require is that these functions be *monotonic*: if $t \sqsubseteq s$ is supposed to mean that s extends t, then if p is true (false) in t, it should be true (false) in s.
TL: I see where we're going now. Let's see if I've got it right: we can extend valuations to complex sentences using the following clauses:

($\neg_\mathbf{1}$) $t \vDash_1 \neg\phi$ iff $t \vDash_0 \phi$;
($\neg_\mathbf{0}$) $t \vDash_0 \neg\phi$ iff $t \vDash_1 \phi$;

[11] It would be better to speak of commitment/anti-commitment or some other more theory-centric notion rather than truth. But it's so much easier to think and speak in truth terms that we will do that instead.

[12] This makes things 'easier' only in a very attenuated sense. In particular, by use of this device, we don't have to detour through an explanation of the so-called Routley-star, which would distract us from the primary purpose of this paper.

($\wedge_\mathbf{1}$) $t \vDash_1 \phi \wedge \psi$ iff $t \vDash_1 \phi$ and $t \vDash \psi$;
($\wedge_\mathbf{0}$) $t \vDash_0 \phi \wedge \psi$ iff $t \vDash_0 \phi$ or $t \vDash_0 \psi$;
($\rightarrow_\mathbf{1}$) $t \vDash_1 \phi \rightarrow \psi$ iff ...

Hang on wait a minute. I can't quite see how ($\rightarrow_\mathbf{1}$) and ($\rightarrow_\mathbf{0}$) should go.
SL: Indeed! This is where things get a bit tricky. Also tricky: we need not only clauses for entailments, but also clauses for both of the bunching operations before we're done!
TL: Of course! Let's see if I can figure out some of these. Hmmm... well, E-Bunches were supposed to be formed by just lumping some things together. So intuitively, an E-bunch should be true just when each of its members is true and false when at least one of its members is. So we ought to have these:

($\mathbf{E_1}$) $t \vDash_1 \{X_i\}_{i \in I}$ iff for all $i \in I$, $t \vDash_1 X_i$;
($\mathbf{E_0}$) $t \vDash_0 \{X_i\}_{i \in I}$ iff for some $i \in I$, $t \vDash_0 X_i$.

Altogether that leaves four clauses I still need some help with: ($\rightarrow_\mathbf{1}$) and ($\rightarrow_\mathbf{0}$) and ($\mathbf{I_1}$) and ($\mathbf{I_0}$)
SL: Two of these are, once you see them, not too bad. Here they are:

($\rightarrow_\mathbf{1}$) $t \vDash_1 \phi \rightarrow \psi$ iff for all u, (i) if $u \vDash_1 \phi$ then $t \circ u \vDash_1 \psi$ and (ii) if $u \vDash_0 \psi$, then $t \circ u \vDash_0 \phi$;
($\mathbf{I_1}$) $t \vDash_1 X; Y$ iff there are u and v so that $u \circ v \sqsubseteq t$ and $u \vDash_1 X$ and $v \vDash_1 Y$.

TL: Let's see if I can make sense of these. ($\rightarrow_\mathbf{1}$) says that a theory makes an entailment true just when, every time we apply it to a theory that makes its antecedent true we end up with a theory that makes its consequent true and every time we apply it to a theory that makes its consequent false we end up with a theory that makes its antecedent false.
SL: Correct!
TL: All that seems fine. ($\mathbf{I_1}$) is a bit trickier, but I think I can see what's going on there too. It's probably important to recognize that both '$\circ$' and ';' are analogues of application—the former being the semantic analogue of application and the latter being the proof-theoretic analogue of application. So if $u \vDash_1 X$ and $v \vDash_1 Y$, then the application of u to v ($u \circ v$) should make true the application of X to Y ($X; Y$). Thus clearly $u \circ v \vDash_1 X; Y$. But ($\mathbf{I_1}$) says not only that $u \circ v$ will do this, but also that anything extending $u \circ v$ will, and that's puzzling me a bit.
SL: Here's a hint: a fairly straightforward induction shows the monotonicity we demanded for atomic formulas can be extended upwards.
TL: Oh excellent! So then yes, given monotonicity, all and only those theories t that extend something of the form $u \circ v$ where $u \vDash X$ and $v \vDash Y$ will make true $X; Y$. What fun! I think I'm on board.
SL: I'm impressed—you're picking things up quite quickly! The remaining two clauses are fiddly. There are explanations of them in [8], but there's really no great reason for us to dwell on them, so let's just let them pass in silence for now. I grant that doing this feels at least a little bit icky. But it helps to recall again that I'm not aiming to do much by presenting this semantics. All I'm trying to do is demonstrate that we *can* build a semantic theory that captures the intuitions we have so far. You may think the details here aren't quite right, but that's ok with me—we can meet up later to work out the details. What matters is that you see the framework can be made to work. Also don't lose track of what I'm trying to convince you of: nothing more than that the universal theory building toolkit is substructural. In particular, I'm *not* interested in arguing that this or that *particular* substructural theory is the universal theory building toolkit.
TL: You're right, you're right. ($\rightarrow_\mathbf{0}$) and ($\mathbf{I_0}$) are small fries. What we're *supposed* to be talking about are the inferences you labeled Rule 1, Rule 2, and Rule 3. Before we deal

with those, I suppose you'll need to tell me what validity amounts to, though I suspect I know what you'll say.
SL: It's not much of a surprise—a formula is valid in a model just when it's true at ℓ in that model. A formula is valid simpliciter just when it's valid in every model.

7. Application and Containment

TL: That's what I thought. Let's return now to the matter of the Rules. I can't help but think that their acceptability is connected somehow to the worries I raised a moment ago about conditions on the application operation that you've used '$\circ$' to refer to in your semantics.
SL: Such rules are indeed intimately connected to the behavior of the $\circ$ operation, so I'm not surprised to hear you say that. Still, I'd like to hear you say a bit more about your worries before we go down that road.
TL: I think what's bothering me is something like this: in the semantics you gave above, you allow for $t \circ t$ to be something other than t. And, I take it from what you say that this isn't an accidental feature of the semantics, but an intentional feature.
SL: Correct! In fact, the semantics allows both for $t \circ t \not\subseteq t$ and $t \not\subseteq t \circ t$. Intuitively, this corresponds to the fact that there are theories that *grow* when applied to themselves and to the fact that there are theories that *shrink* when applied to themselves.
TL: I'm trying to understand how this could possibly be the case, and I think I see what's bothering me. Let's say a theory is nonidempotent when it *grows* when applied to itself.
SL: Nonidempotent is exactly the sort of ugly word I'm happy to take in; go on.
TL: Well here's the thing: suppose th is a nonidempotent theory, and write th^2 for the result of applying th to itself. Since th is nonidempotent, $th^2 \not\subseteq th$. But th^2 is, given what we've said before, $\{\psi : \phi \to \psi \in th \text{ and } \phi \in th\}$. Thus, since $th^2 \not\subseteq th$, for some ϕ and ψ we will have all of the following:

- $\phi \to \psi \in th$, and
- $\phi \in th$, but
- $\psi \notin th$.

SL: That's correct!
TL: Well I think it goes without saying that any such 'theory' is degenerate and ought to be rejected—if you're not closed under modus ponens, you just not a theory at all!
SL: I completely disagree; there's nothing at all degenerate about such a th! Remember: entailments determine a theory's *background* behavior. We spelled this out in the semantics above by saying that th makes true $\phi \to \psi$ iff any time we *apply* th to a theory where thus-and-so happens, such-and-such occurs. The details aren't super important. What matters is that $\phi \to \psi$ and ϕ don't interact when minding their own business inside th. To get them to play together we have to *apply* th to th. So if you could build a theory th so that $\phi \to \psi \in th$, $\phi \in th$, but $\psi \notin th \circ th$, then *that* would be a problem. But what you've pointed out here as a bug is actually just a feature.
TL: I'm not convinced the bug isn't a bug. The problem—and now I'm finally going to say what I've wanted to say all along—is that you've misunderstood what it means to take th as a background theory. All it means to take th as a background theory is to restrict our attention to situations compatible with th. If you want to think about it model-theoretically, this means all we need to do is restrict our attention to models that verify every member of th. So when we talk about, e.g. 'the th-consequences of f' what we're interested in is really just the consequences of f that are compatible with th, and this just means we're interested in $th \cup f$. Ergo applying th to f is just forming the union of th and f.

SL: Well…
TL: Let me finish! You've already accepted that applying a theory committed to $\phi \to \psi$ to a theory committed to ϕ results in a theory committed to ψ. So if *th* is committed to $\phi \to \psi$ and to ϕ then the theory that results from applying *th* to *th* is committed to ψ. But I just pointed out that the theory that results from applying *th* to *th* is just the theory $th \cup th$, which is clearly just *th*. So *th* is, if committed to $\phi \to \psi$ and to ϕ, already committed to ψ.
SL: Your objection strikes right to the heart of the matter. As you probably expect, I reject your claim that application is set union. But this follows from what I've already said: we cannot accept any account of the universal theory building theory that forces us to accept Rules 1-3.
TL: Wait just a minute, I'm confused—I don't see the connection between my set-theoretic reading of application and these rules. I meant the objection to just be an objection to what I took to be a core part of your semantics.
SL: As I mentioned a moment ago, features of the application operation are tightly connected to these kinds of rules. So it shouldn't be too surprising that your proposal to read application as set union leads us to accept such rules. Nonetheless, let's have a look to see how this plays out.
TL: Great! Let's focus on the rule that says we can infer ψ from $(\phi \to \psi) \wedge \phi$.
SL: Alright then. Suppose we've adopted your set-theoretic reading of application, and suppose also that we have on hand a theory that contains $(\phi \to \psi) \wedge \phi$. In the language we introduced a bit ago, that means we're working with some t such that $t \vDash_1 (\phi \to \psi) \wedge \phi$. Then $t \vDash_1 \phi \to \psi$ and $t \vDash_1 \phi$. Thus $t \circ t \vDash_1 \psi$.
TL: Ah I see now: if '$\circ$' names set union, then $t \circ t$ is the same thing as t. So then since $t \circ t \vDash_1 \psi$, we also have that $t \vDash_1 \psi$. Very good! I take it that it's also the case that were I to adopt the set-theoretic interpretation, I would be forced to accept the other two rules? And the reasons are similar?
SL: Indeed. The details are slightly more complex, but the basic principle is the same.
TL: Ok, we've established a connection between my claim that application is set union and the rules you're challenging. Are you finally willing to tell me why you have a problem with these rules?
SL: I am. But I'm warning you: my reason is quite boring.
TL: I have braced myself for boredom. Please proceed.

8. Theory Building Again

SL: Very well. According to Rule 1 (for example), every theory that contains $\phi \to (\phi \to \psi)$ also contains $\phi \to \psi$. But we can build theories do contain $\phi \to (\phi \to \psi)$ but do not contain $\phi \to \psi$. So we have to reject Rule 1. Thus in particular, we have to reject that application is set union. For similar reasons, we also have to reject the other two rules.
TL: I'm afraid you've left me a rather easy route out: I only need to reject that you can build such theories. Of course, I accept that you can construct *sets of sentences* that contain $\phi \to (\phi \to \psi)$ without containing $\phi \to \psi$. That's easy. But you can also construct sets of sentences that contain ϕ and contain ψ without containing $\phi \wedge \psi$. Just as you'd agree that the goal of building a *theory* rules out the latter, you must also agree that it also rules out the former.
SL: I agree with you: anything that counts as a theory will, if it contains ϕ and contains ψ also contain $\phi \wedge \psi$. Why do you think the same thing is true of $\phi \to (\phi \to \psi)$ and $\phi \to \psi$?

TL: Well let's see. It's not something I've ever had to justify before, but I guess it's something like this: when we set out to build a theory, our aim is to build *as-complete-as-possible* theories.[13] But note that whenever $\phi \to (\phi \to \psi)$ is true $\phi \to \psi$ is also true. So there's no harm in adding $\phi \to \psi$ to our theory if we've already added $\phi \to (\phi \to \psi)$ to it, since we cannot find a counterexample to one that isn't a counterexample to the other. And since there's no harm in adding it, there *is* harm in keeping it out, since doing so keeps out theory from being as complete as possible. Ergo we should concern ourselves with idempotent theories only.[14]

SL: This is a sophisticated objection. My knee-jerk response was to say that in fact there *are* counterexamples to $\phi \to \psi$ that are not counterexamples to $\phi \to (\phi \to \psi)$, and to then just build one using the semantics we built above.

But I can see that this isn't likely to move you. You'd simply say that this shows the semantics is wrong—it was, after all designed with the express intent of capturing the behavior of theories under application. To appeal to it in order to justify my account of how theories behave under application is thus flatly circular. So I'll need a different argument.

9. Curry

*There's a long moment of silence while **SL** gathers his thoughts while staring just to the left of the camera. Just as it starts to get awkward, he looks at the camera and smiles.*

SL: Ok, I think I see where to go. I'll start on your territory, where Rules 1-3 are all allowed.

TL: How kind of you.

SL: We'll see. Here's a fun fact: since we're accepting Rule 2, all theories are idempotent.

TL: Interesting! Let's see if I can see why this would be. Hmmm. . . well, let's suppose ψ is in the application of t to itself. Then there will be a $\phi \in t$ so that $\phi \to \psi$ is also in t. But then since $\phi \to \psi$ and ϕ are both in t, so is $(\phi \to \psi) \wedge \phi$—you've already admitted this rule. So by Rule 2, ψ is in t. And there we have it! Everything in the application of t to t is in t, so t is idempotent.

SL: Excellently done! Now suppose our theory building concerns truth in languages with enough arithmetic around to do some diagonalization. We won't need to get too technical. All we need is that we can build a 'Curry sentence'; that is, some sentence c so that for some sentence U that we're not willing to accept, the following hold:

$$\text{C1: } c \to (T\ulcorner c \urcorner \to U) \qquad \text{C2: } (T\ulcorner c \urcorner \to U) \to c$$

Since C1 and C2 are true, we might add them to our theory. Maybe we also observe that T1 and T2 are true and add them to our theory:

$$\text{T1: } T\ulcorner c \urcorner \to c \qquad \text{T2: } c \to T\ulcorner c \urcorner$$

TL: Well now I'm suspicious, but go on.[15]

SL: Thanks. We'll have time to deal with your suspicions later. I claim that if we include C1, C2, T1, and T2 in our theory, and require that the theory be idempotent, then it will also contain U. But we said above that U was something we were unwilling to accept.

TL: I'd like to see this argument!

SL: It's fairly short, so it's worth being quasi-formal about it. Here, let me write it:

---

[13] This is an important consideration in [1], for example.

[14] Eric Stei presented a forceful version of this line of argument in response to a talk I gave on related material at the North Carolina Philosophical Society in 2018.

[15] Thanks to Ben Caplan for raising the suspicion this eventually leads to.

***SL** scrolls to a blank page in the shared screen and, slowly and laboriously, scrawls the following:*

(1) By assumption, $T\ulcorner c\urcorner \to c$ is in our theory.
(2) So by Rule 3 applied to (1), $(c \to (T\ulcorner c\urcorner \to U)) \to (T\ulcorner c\urcorner \to (T\ulcorner c\urcorner \to U))$ is in our theory.
(3) By assumption, $c \to (T\ulcorner c\urcorner \to U)$ is in our theory.
(4) Since our theory is idempotent, we conclude from (2) and (3) that $T\ulcorner c\urcorner \to (T\ulcorner c\urcorner \to U)$ is in our theory.
(5) Rule 1 applied to (4) tells us that $T\ulcorner c\urcorner \to U$ is in our theory.
(6) By assumption, $(T\ulcorner c\urcorner \to U) \to c$ is in our theory.
(7) Since our theory is idempotent, we conclude from (5) and (6) that c is in our theory.
(8) By assumption, $c \to T\ulcorner c\urcorner$ is in our theory.
(9) Thus, again appealing to the fact that our theory is idempotent, we conclude from (7) and (8) that $T\ulcorner c\urcorner$ is in our theory.
(10) Thus, by one final appeal to the theory being idempotent, we conclude from (5) and (9) that U is in our theory.[16]

10. Concluding Thoughts

TL: Slick! You've shown that if we accept rules 1-3 and there's anything at all that we're unwilling to accept, then we're in trouble if we try to go about building a theory that contains C1, C2, T1, and T2.
SL: Exactly. So, since we can build theories containing these sentences without getting into trouble, we can build theories that reject some of these rules.
TL: I think you're being too quick here, friend. This gets back to my suspicion from before. The right conclusion to draw from this isn't that the universal theory building toolkit can't accept Rules 1-3, but that we have to avoid saying awful things like C1 and C2 (or at least, we have to avoid saying such things if we also want to say things like T1 and T2).
SL: Let me get this straight: your solution is to save the *universal* theory building theory by *not* thinking about certain theories. I don't understand how that can possibly be correct.
TL: Well when you put it like *that* it sounds pretty bad, I'll admit. Perhaps what I mean to say is this: adding truth predicates and diagonalization to the mix *always* causes problems. This seems to give strong prima facie evidence for their being the type of things best handled separately using very thick gloves, not barehandedly as you've done here.

Less metaphorically, what your argument shows is that these concepts are unapt for theory-work and should either be abandoned or replaced.[17]
SL: I have two replies to this. First, truth and related notions feature prominently in our semantic theories, in metaphysics, and in a host of other places. If the universal theory building toolkit trivializes all of theories containing such notions that are also complex enough to do some diagonalization, then it seems we just cannot do non-trivial work in semantics, metaphysics, and the like.

Thus if we think we *can* do e.g. semantics or metaphysics, then its incumbent on us to think through what that says about the structure of the universal theory building toolkit.

Second, I will grant that *if* we demand all theories to obey Rules 1-3, *then* of course we cannot build theories containing C1, C2, T1, and T2. So there are two options here: drop our demand that all theories obey Rules 1-3 or drop theories containing C1, C2, T1, and

[16] Andrew Tedder has helpfully pointed out that this is quite similar to the derivation in [5].
[17] See [12] for an illuminating discussion of the latter option.

T2. I suppose what I want to claim is that the former option is the more intelligible and appealing option.

TL: I think I see what you're saying, even if I disagree. Maybe we could say it like this: if we began naïvely in our search for the universal theory building toolkit, then we wouldn't have any reason to reject theories containing C1, C2, T1, and T2 or to demand that theories obey Rules 1-3. We find the latter option 'intuitive' largely because of our extensive training in traditional logic. But you've provided an alternative way of understanding theories and theory building that plays into the naïve intuitions and is otherwise quite compelling as well. It's thus on the traditionalist at this point to provide a defense of her position.[18] This is a nice line of argumentation, but it leaves one question open: what evidence can you actually marshal in defense of the view of theories you've provided? I think you've made clear that if we accept the broad outlines of your account—e.g. if we accept that there are two theory building operations; that they behave in basically the way you outline; that theories are sets of sentences; etc.—then we should accept that the universal theory building theory is substructural. What can you say in defense of *those* hypotheses, though?

SL: That's a very serious question. I tend to think of what I'm doing in a broadly Carnapian way—what I'm presenting here is a *theory* about what sorts of things theories are and about how theory building goes. I'm then appealing to pragmatic considerations to urge adopting my theory building theory over its rivals. The kinds of pragmatic considerations I'm aiming at include both the paradox-avoidance bits just highlighted as well as the fact that my theory allows us to go on talking about theories in many of the ways we customarily do. I don't expect we have time for me to say more about this now, but, generally speaking, that's the sort of answer I'd tend to give to your question.

TL: I suppose that will have to do for now. There's one other objection I'd like to raise: one might, it seems to me, accept that there are theories that contain C1, C2, T1, and T2, but also hold that messing about with such things isn't the domain of *logic*. Logic is concerned with more pristine matters where behavior like nonselfclosure doesn't arise.

SL: It's true that Beall (e.g.) has argued that the there is an important and philosophically central sense of the word 'logic' on which logic *just is* the universal theory building toolkit.[19] I personally don't see a need to take a stand on how the universal theory building toolkit and logic are related. This is because I take it to be clear that whether the universal theory building toolkit is identical to logic or not, the theory itself is of central philosophical importance. I've scrupulously avoided the claim that what I'm doing is logic. If you'd like, I'm happy to give you the word 'logic' to use for the game we play when we restrict our attention to idempotent theories. Logic, in that case, is not the *universal* theory building toolkit. It's a special-case theory building toolkit. So anytime you're out and about and sit down to build a theory, you need to check to see whether you can use logic. That's not *usually* how we expect logic to work, but maybe you can make a case for using the word this way. In any event, I'll be over here if you need me, exploring what we can use to build theories all the time.

At the end of this speech, the connection suddenly and irreparably breaks. ***SL*** *and* ***TL*** *are forced, as happens so often, to perform the usual goodbyes and see-you-soons and via email. But they return to their quarantined lives refreshed by this pleasant interaction.*

---

[18] I have never encountered a traditional logician who is this conciliatory to the substructural position. Perhaps the best way to read this part of the dialogue, then, is not as a record of some possible interaction between a traditional logician and a substructural logician, but rather as a model for how such interactions could go, were dialogue between these two sides to be more fruitful than it in fact tends to be.

[19] See [1] and [2].

References

[1] Jc Beall. There is no logical negation: True, false, both, and neither. *Australasian Journal of Logic*, 14(1):Article no. 1, 2017.

[2] Jc Beall. The simple argument for subclassical logic. *Philosophical Issues*, 28(1):30–54, 2018.

[3] Robert Brandom. *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard university press, 1998.

[4] Robert Brandom. *Articulating reasons: An introduction to inferentialism*. Harvard University Press, 2009.

[5] J Michael Dunn, Robert K Meyer, and Richard Routley. Curry's paradox. *Analysis*, 39(3):124–128, 1979.

[6] Kit Fine. Models for entailment. *Journal of Philosophical Logic*, 3(4):347–372, 1974.

[7] Steve Giambrone. TW+ and RW+ are decidable. *Journal of Philosophical Logic*, 14(3):235–254, 1985.

[8] Shay Allen Logan. Deep fried logic. *Erkenntnis*, pages 1–30, 2020.

[9] Robert K Meyer. The consistency of arithmetic. *Typescript*, 1975.

[10] Robert K Meyer and Richard Routley. Algebraic analysis of entailment I. *Logique et analyse*, 15(59/60):407–428, 1972.

[11] Stephen Read. *Relevant Logic: A Philosophical Examination of Inference*. Blackwell, 1988.

[12] Kevin Scharp. *Replacing truth*. OUP Oxford, 2013.

[13] John Slaney. A general logic. *Australasian Journal of Philosophy*, 68(1):74–88, 1990.

[14] Alasdair Urquhart. Semantics for relevant logics. *The Journal of Symbolic Logic*, 37(1):159–169, 1972.