The *I* in Logic*

Gillian Russell

This is the draft of a paper written for the symposium following the presentation of the Rolf Schock Prize to David Kaplan at the Swedish Academy of Sciences in Stockholm in October of 2022. It will appear in a special issue of the journal *Theoria* along with the other papers from the symposium. Please cite official version once available.

Autobiographical Prelude

Kaplan's *Demonstratives* had a big influence on me. I first read it when I was grad student in philosophy at Princeton, getting ready for the so-called "Generals exams." Generals aren't all that general; you and your professors come up with list of books and papers on the "general area" in which you plan to write your dissertation and you spend the semester reading it all, and then there are written and oral exams. *Demonstratives* was on my list (almost certainly thanks to Scott Soames) and I still remember where I was when I read it. I'd been feeling overwhelmed by the scale of a semester-long task followed by exams that students can—and did—fail. But that day I was determined to give it my best shot, and I sat down in back of the Campus Center coffee shop with a notebook, the right kind of pen, coffee—of course—and my first xeroxed copy of "Demonstratives" (there would be several more steadily disintegrating photocopies, until I finally bought my own *Themes from Kaplan* after graduating to being the kind of academic who has research funds.) It must have taken me more than a day to read the entire thing, but I broke the back of it that first day and perhaps the way to say what I need to it to say that I found it absolutely *beautiful*. I could try to explain why—some of it has to do with the "metaphysical picture" and the way propositions, direct reference, indexicals, and rigid designation hang together. Some of it is the formal system, and the way that connects the picture up with a broader set of entrenched ideas. Some of it is—as Kaplan says himself—about the revelation that comes from pursuing obvious ideas to interesting conclusions. But honestly my faith in my

*This paper benefited from discussion at the Symposium following the presentation of the Schock Prize to David Kaplan at the Swedish Academy of Sciences in Stockholm in October 2022. My thanks especially to Andreas Stokke and Joseph Almog for helpful comments at the Symposium, and to Peter Pagin and Dag Westerståhl for detailed comments on a written draft.

own ability to explain why I found it beautiful is not terribly high. Suffice it to say: I thought it was beautiful and it made a big contribution to my faith that worthwhile work was *possible* in philosophy. I wasn't very sure what I was supposed to be doing but *Demonstratives* gave me a sense of what I wanted to aim for—I wanted to try to do work like *that*. I once overheard a couple of students speaking contemptuously about it—actually an almost ubiquitous part of my grad school experience was hearing really smart people speak contemptuously of work in philosophy—Aristotle, Nussbaum, Kant, Tarski, nobody was immune, *especially* whoever that week's speaker had been—but hearing them diss *Demonstratives* was, I think, the first time I heard such talk and, instead of feeling dumb for not seeing why the target was so *obviously* wrong, I thought (with some surprise) "hey, those guys *just don't get it*." And this is a first faltering step to such talk losing some of its power.

When the official part of this paper gets going in the next section I'll claim that *Demonstratives* has been very influential in philosophy, and it wasn't hard to source quotations to back up that claim. But *influential* in that sense "is other people." The reason I wanted to write this paper is that "Demonstratives" had a huge influence on me personally—on my work, and on my faith in philosophy. And for that I just wanted to say: thanks.

1 Introduction

"Demonstratives" and its successor "Afterthoughts" (Kaplan, 1989a,b) have been enormously influential:

> Understanding of demonstrative semantics grew by a quantum leap in David Kaplan's remarkable work, especially in his masterpiece "Demonstratives" (Salmon, 2002, 497)

> By far the most influential theory of the meaning and logic of indexicals is due to David Kaplan. Almost all work in the philosophy of language (and most work in linguistics) on indexicals and demonstratives since Kaplan's seminal essay "Demonstratives" has been a development of or response to Kaplan's theory. (Georgi, 2022)

> Kaplan's theory of indexicals is highly influential and serves as a starting point for much current theorizing about indexicals. (Braun, 2017)

> An elegant formal apparatus has been devised by David Kaplan... probably the single most influential contribution to our topic. (Forbes, 1989, 464)

> In the classic analysis of indexicals, namely David Kaplan's *Demonstratives*... (Predelli, 2008, 57)

> *Demonstratives*—one of the most frequently cited unpublications . . . (Berg, 1991, 92)

That said, I think "Demonstratives" deserves more uptake than it generally gets *in logic*. In his preface Kaplan wrote "I now think that...the most important part and certainly the most convincing part of my theory is just the logic of demonstratives itself." (487)[1] I agree about the importance of the logic, and here I want to focus on it as one of three important things from "Demonstratives," the other two being the simple background metaphysical picture of the way indexicals work, and the consequences these two things (the picture and the logic) have for our conception of logical properties like logical truth and consequence.

I can think of two ways to highlight the significance of the logic of demonstratives. The first is to look at what things would be like in its absence—and I'll make a start on that in the next section. The other is to give an example of what we can do now that we do have it—an application—and the second half of this paper will pursue one such application, before using it as the basis for some concluding remarks about the ways in which a formal logic can be useful.

2 No *I* in logic

One way to appreciate the importance of indexical logic is by looking at where we would be without it. We can get some insight into that by looking at what philosophers and logicians tended to say about indexicals and validity before *Demonstratives*.[2] The strategies for dealing with indexicals that follow fall into four overall groups:

1. there is no formal logic for indexicals (logic is restricted to the languages of mathematics and other serious sciences, whose languages don't contain them.)
2. informal paraphrase
3. fix a context
4. formal paraphrase

P. F. Strawson is one author who thinks there is no formal logic for indexicals. He argues that there can be no logic for natural language *because* it

[1] Kaplan calls his formal system *LD*, for *Logic of Demonstratives*, but that might be misleading, since LD is also (and perhaps in the first place) a logic for *indexicals* like *I*, *here*, and *now*. I have occasionally met philosophers who claim that LD is "not a formal system" on the grounds that it is not an axiomatic proof system. I won't assume such a restricted understanding of "formal system" here, though I'll usually use the word *logic* instead. I take it that there is a perfectly good reading of *logic* on which one can be specified model-theoretically.

[2] Some of the authors below—like Strawson, Quine, and Jeffrey—were literally writing before the publication of *Demonstratives* in the temporal sense of *before*. Others were (for reasons practical or theoretical) simply ignoring it and so we might instead think these as prior to *Demonstratives* conceptually, pedagogically, or on the great path of philosophical progress.

contains indexicals.[3] The heart of his view is that it is important to distinguish expression types from uses of expressions in contexts. Types are associated with rules and meanings that speakers learn when they learn the language (also called *standing meaning*, or *character*.) These meanings are independent of context and this is the level at which we get rules concerning entailment. But Strawson points out that *referring* expressions (most obviously indexicals like I, but also, famously, definite descriptions) only get their referents once they are used in a context. Successful reference is required for a sentence containing a term to have a truth-value, and so sentence-types won't generally have truth-values. That is a problem if formal logic is (as I think Strawson takes it to be) about truth-preservation over sentence types.[4] He allows that there are some special expression types (perhaps most commonly in the domains of mathematics or the necessary) where there is no variation of reference with use. But natural language as a whole contains many expressions whose referent varies between uses—the paradigm being the indexical I—and for this reason there can be no formal logic of natural language. The domain of formal logic then is mathematics and perhaps science, where language is indexical-free, and reference non-contingent and timeless.

This strategy of focusing on indexical-free language is one way to avoid arguments containing indexicals.[5] But philosophers—if not mathematicians—should be wary of this because there are important *philosophical* arguments containing indexicals. The most prominent is certainly Descartes' *cogito:*[6]

[3](Strawson, 1950, 1952) See also Radulescu (2015) who writes: "One of the central arguments in Strawson ... against the possibility of a formal semantics of natural language and of an associated logic was that indexicals cannot be dealt with formally." (1839) Strawson's case specifically targets formal logic, and in this paper I'll be using *logic* with this meaning, since I'm interested in the significance of LD—a formal logic for indexicals.

[4] "From the logician's point of view, the ideal type of sentence is one of which the meaning is entirely given by entailment-rules; that is, it is one from which the referring element is absent altogether; that is, roughly, it is one of which it is true that if its utterance at any time, at any place, by any speaker, results in a true statement, then its utterance at any other time, at any other place, by any other speaker, results in a true statement. Almost the only types of contingent sentence (i.e., sentence the utterance of which would result in a contingent statement) which seem able fully to realize this ideal are positively and negatively existential sentences, of which some forms are studied by the predicative calculus, or sentences compounded of these." (Strawson, 1952, 214)

[5]See also (Jeffrey, 1967, 5–6)

[6]The argument is standardly attributed to Descartes, though it doesn't appear in the *Meditations* in this famous form. In the second Meditation he puts it this way: "I have convinced myself that there is absolutely nothing in the world, no sky, no earth, no minds, no bodies. Does it now follow that I too do not exist? No: if I convinced myself of something then I certainly existed. But there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. In that case I too undoubtedly exist, if he is deceiving me; and let him deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something. So after considering everything very thoroughly, I must finally conclude that this proposition, *I am, I exist*, is necessarily true whenever it is put forward by me or conceived in my mind." (Cottingham et al., 1984, 2:16f) (Newman, 2019) (There is some irony to the last clause, given that Kaplan helped us clarify that *I am* does not become necessary when *I think* is true. So this turns out to be one more way in which reading Kaplan helps us to process Descartes.)

$$\frac{\text{I think.}}{\text{I am.}}$$

Some philosophers joke about the cogito in discussions of indexical logic, but here I propose to use it seriously as a test for pre-Kaplanian strategies for dealing with indexicals. As I see it, the cogito has four features that we need to take account of. First, it is philosophically important, both intuitively and historically. Second, it has a premise which seems strikingly epistemically accessible: even when I feel most threatened by global skepticism, even after I am forced to concede that much else has been made doubtful, it still seems clear that I am thinking (or doubting). Third, the argument has the appearance of validity, the property central to logical theory. And fourth, the argument's conclusion is significant: that is, it *matters* (to me) whether or not I exist, and presumably it matters to each reader whether nor not *they* exist. None of these four features of the argument is beyond challenge (if anything in philosophy ever is) but logical consequence is about capturing the intuitive conception of validity, and this at least appears to be an important example. Logical theory needs to either explain it, or explain away the appearances.

Strawson's approach declines to take on the challenge, arguing instead that it is impossible, since logic cannot handle indexicals. If there *is* a positive account that completes the task, that will show, first, that it was never impossible, and second, such an approach would be superior to one which restricts logic to non-indexical mathematical and scientific discourse, since it will unify the theory of validity for the language of mathematics and science with that of philosophy.

Most authors are not as explicit as Strawson about logic's phobia of indexicals. Indeed, it is surprisingly common for introductory textbooks in logic to bring up indexicals in the first few pages. Goldfarb's *Deductive Logic* discusses the sentences "I am myopic" and "Seattle is far from here" (4–5), Jeffrey's *Formal Logic: it's Scope and Limits* uses "Of all the men of this time whom I have known..." (4, quoting Socrates via Plato) and Quine's *Methods of Logic* notes that "The pronoun "I" changes its reference with every speaker; "here" changes its reference with every significant movement through space; and "now" changes its reference every time it is uttered." (1).[7] But these examples are not a prelude to a logic *for* indexicals, since the authors hold, as Jeffrey puts it:

> Resolution of vagueness, ambiguity and context-dependence is a preliminary to formal logic, not a part of it. (Jeffrey, 1967, 7)

They generally go on to suggest various methods for eliminating the context-dependence introduced by indexicals. One method (no. 2 in the list above) is paraphrase. Goldfarb replaces "Seattle is far from here" with "Seattle is far from Philadelphia" and Jeffrey transforms "Of all the men of his time whom I have known, he was the wisest and the justest and best" into "all the men

[7](Quine, 1950), (Goldfarb, 2003), and (Jeffrey, 1967) These textbooks were chosen on grounds of being especially prominent, rather than because they present especially egregious examples. Textbook authors may have pragmatic reasons to avoid going into the details of an indexical logic.

of Socrates' time whom Phaedo knew, Socrates was the wisest and the justest and best." Goldfarb notes that where a sentence contains tensed verbs, explicit mention of the time must be inserted. He gives the example of paraphrasing "Maria Callas will sing at the Metropolitan Opera" to "Maria Callas sings at the Metropolitan Opera after May 6th 1963." (5)

If we apply this strategy to the *cogito*, we might arrive at:

$$\frac{\text{The author of } \textit{The Meditations} \text{ thinks at 10am on 1st January 1641.}}{\text{The author of } \textit{The Meditations} \text{ exists at 10am on 1st January 1641.}}$$

This is valid, but the transformation is cumbersome, and it is hard to know exactly which words to use. (Why 10am and not 9am?) But the main problem is that this paraphrasing destroys philosophically important properties of the argument. I have no special epistemic access to the truth of the new premise; for all I know the author of *The Meditations* was unconscious at that time. Similarly, the new conclusion does not matter to me in the same personal way: the original argument was something any one of us could use to establish *our own* existence. I was promised *I exist*; it seems a poor substitute to be left with *Descartes used to exist.*

A third strategy suggested in Goldfarb (5–6) is to *fix a context* for the entire argument, relative to which any indexicals will have a constant referent. We might suppose, for example, that the unparaphrased cogito is uttered by Descartes at 10am on 1st January 1641. Then the *I* in both the premise and conclusion refers to Descartes, and the premise tells us that Descartes thinks at that time, and the conclusion that Descartes exists at that time. This is less cumbersome than the last strategy and it requires less creativity. But, like informal paraphrase, it jettisons some of the argument's philosophically important features. We can relativise the argument to different particular contexts, but in nearly all cases that will cause it to become a deduction of something less interesting to us, on the basis of something we have less reason to believe. So this strategy seems to miss something of what makes the argument important. But it also misses a key general pattern: we don't need to relativise to a speaker and time before assessing for validity, because *whoever* is speaking, at whatever time, truth is preserved: this is surely obvious to ordinary readers of the argument, who recognise that the argument preserves truth in Descartes' context, as well as in their own as they read and think about it, and in contexts in which others—their students, readers and interlocutors—take themselves and their own times to fix the context.[8]

A final pre-Kaplanian strategy suggested in introductory logic texts is *formal paraphrase.* (Jeffrey, 1967, 6–7) This time we replace context-sensitive expressions with non-context-sensitive, non-logical constants, as a part of the process of translating the argument into the language of some formal logic. For example, we might translate the *cogito* as:

[8]It is, however, as one referee notes, important to keep the context the same over the course of the argument.

$$\frac{a \text{ thinks.}}{a \text{ exists.}}$$

and finally as:

$$\frac{Ta}{\exists x(x = a)}$$

We might go on to note that on some conceptions of logical consequence, to say that this argument form is valid is to say that however we interpret the non-logical expressions—i.e., no matter that T and a mean—the argument will have a true conclusion if it has a true premise. So now—unlike with the "fixing a context" strategy—we do seem to have done a better job of grasping the general picture: no matter what a refers to, if a "Ts" (does or is something—anything), then a exists.

Still, we learned from Kaplan that this strategy—though an improvement—misses something key to the logic of indexicals. To see what that is, it helps to think a bit about logical truth, and a bit about meaning.

3 Indexicals and the Logical Properties

Here is a picture of how language—a least some bits of it—works. Sentences are composed of words. A word is a string of characters with a meaning and the meanings of sentences are composed of the meanings of the words they contain. Declarative sentences say things about the way the world is. If what the sentence says is true, then the sentence itself is true—it inherits truth from what it says. Otherwise the sentence is false. Call the thing the sentence says the *proposition* it expresses, and (following Kaplan) that kind of meaning (propositional) *content*. Then words have propositional content and the propositions expressed by sentences are composed of the contents of the words in the sentences.

Many sentences are true in some circumstances, false in others. The logical truths are supposed to be special: they are true (to put it a bit vaguely for now) come what may. To put it this way is to hint at a certain relationship between logical truth and necessity: one might expect logical truths to express necessary propositions, propositions which are true come what may, i.e. true in every possible world. We might even be tempted by the view that a sentence is a logical truth iff it expresses a necessary proposition. But the relationship between sentences that are logical truths and propositions that are necessary rather depends on the relationship between sentences and propositions, and getting the former right requires care with the latter.

If sentences stood in a 1-1 relationship to propositions, we could use sentences as a kind of proxy for the propositions they express.[9] Then the only way for a sentence to be true come what may would be by expressing a necessary truth.

[9]Quine complains that many authors simply fail to distinguish them adequately: "Philosophers' tolerance towards propositions has been encouraged partly by ambiguity in the term 'proposition'. The term often is used simply for the sentences themselves, declarative sen-

But sentences do not stand in a 1-1 relationship to propositions, because, first, two different sentences can express the same proposition. *Snow is white* and *Schnee ist weiß*, for example, or *Hesperus is bright* and *Phosphorus is bright*, or Pa and Pb (relative to a model in which $a = b$ is true), or $a = a$ and $a = b$. The sentence-proposition relationship is many-1, rather than 1-1, and so expressing a necessary truth is not sufficient for being a logical truth. For $a = a$ is a logical truth and $a = b$ is not—even if they express the same necessary proposition.

And then consider indexicals, whose defining characteristic is that their propositional content varies with context of use, so that a single sentence containing one can express one proposition in one context, another in another. This complicates the relationship between sentences that are true come what may, and propositions that are. Because now a sentence can express a necessary proposition without being always true itself—for example *I am Gillian* expresses a necessary proposition relative to the context in which I am the agent, but a false one relative to the context where *Andreas* is. Moreover—as Kaplan showed us—a sentence can be true-come-what-may without the proposition that it expresses being necessary. It might even express a contingent proposition in every context—so long as in every context the proposition it expresses is true *in that context*. The sentence *I am here* often functions as a useful illustration of this idea. Said by me in Melbourne it expresses the proposition that Gillian is in Melbourne, and that proposition is true, though not necessary—I could have been in Stockholm. Said by Andreas in Uppsala, the same sentence says that Andreas is in Uppsala, which is true in that context (though again not necessary.) And now it is clear that relative to any context, the sentence is true—it is true come what may—though the propositions it expresses are (usually) non-necessary. Other examples of contingent logical truths in LD include $dthat[\alpha] = \alpha$, *The actual B is the B*, and $Ap \leftrightarrow p$.

There is a high-profile controversy about whether *I am here now* is a genuine example of a contingent logical truth.[10] But ultimately it doesn't matter very much for the thesis that there are contingent logical truths whether that sentence turns out to be one of them. Arguing over whether it is a contingent logical truth is analogous to arguing over whether the law of excluded middle is an ordinary logical truth; even if it isn't, that does little to undermine the thesis that there *are* logical truths, we are merely arguing over the extension. We didn't come to see that there are contingent logical truths by generalising from particular contested instances. It is the Kaplanian picture of indexicality itself that makes space for them; once a sentence can express different propositions in different contexts, it is no longer forced to inherit the robustness of its truth from the proposition it expresses. Propositions are the primary bearers of truth, but not of logical truth.

This point generalises to the other logical properties. Two sentences can be logically equivalent, though one expresses a necessary proposition and the

tences; and then some writers who do use the term for meanings of sentences are careless about the distinction between sentences and their meanings." (Quine, 1986, 2)

[10] See e.g. Vision (1985); Predelli (1998)

other a contingent one. In LD, $A\phi$ and ϕ have this feature for contingent ϕ, as does does the formalisation of *I am here now*, $\mathcal{N}Loc(\mathfrak{i}, \mathfrak{h})$, paired with any non-contingent logical truth, such as $p \vee \neg p$.[11] A fortiori, the premises of a valid argument might be necessary and the conclusion contingent, suggesting that it is not really *necessary* truth-preservation that is the heart of logical consequence, but rather, truth-preservation in virtue of the meanings of the logical expressions.

The pre-Kaplanian strategies that we looked at—restriction to non-indexical language, formal and informal paraphrase, as well as fixing a context—miss these features because they replace indexicals with non-logical expressions or examine them only in one context—in effect eliminating the contextual variation that is their hallmark. But the features of the logical properties that we identified above depend on the meanings of the indexicals. And the feature of Kaplan's logic that allows it to bring this out is that it treats indexicals as logical constants, thus preserving their characters over models. This allows us to identify sentences that have the logical properties they do as a result of containing indexicals.

4 Indexical Logic

To illustrate the above, and as a propaedeutic for looking at applications, we need an indexical logic. LD—the one in "Demonstratives"—is really complicated: a two-sorted quantified modal tense logic with identity and functors even before we add contexts. The one I present here—call it IL, for indexical logic—is somewhat simpler, but what makes it an indexical logic it takes from Kaplan: logical constants that include indexicals, and models that feature contexts (though here just a single context per model) which—together with the characters of indexical expressions—determine the referents of indexical terms and the truth-conditions of sentences containing indexical operators.

4.1 The Language

Terms:

individual variables	x, y, z, etc.
position variables	v_1, v_2, v_3, etc.
individual non-logical constants	a, b, c, ...
position non-logical constants	p, p_1, p_2, ...
an individual logical constant	$\mathfrak{i}$
a position logical constant	$\mathfrak{h}$

There are two kinds of term: one for referring to individuals (i-terms), one for referring to positions (p-terms). Non-logical constants, variables, and logical

[11] As Peter Pagin points out to me, this assumes that validity is truth in all contexts, which is in line both with Kaplan's approach and with (it seems to me) the intuitive classification of logical truths and consequences.

terms thus each come in two varieties. The two logical terms are both indexicals: $\mathfrak{i}$, for the first person pronoun *I*, and $\mathfrak{h}$, for the pure indexical *here*.

Predicates:

non-logical predicates	$A^{\langle m,n\rangle}$, $B^{\langle m,n\rangle}$, $C^{\langle m,n\rangle}$, etc.
special predicate	$Loc^{\langle 1,1\rangle}$, $Ex^{\langle 1,0\rangle}$
logical predicates	$=^{\langle 2,0\rangle}$, $=^{\langle 0,2\rangle}$

Because there are two kinds of term, the arity of a predicate is given by a pair of numbers, $\langle m, n\rangle$, with m the number of i-terms, and n the number of p-terms taken to form a formula. We will often drop these arity-indicating superscripts in the interests of readability. The special predicates *Loc* and *Ex* may be read informally as *is located at* and *exists*.

Connectives and Operators

truth-functional connectives	$\wedge$, $\vee$, $\neg$, $\rightarrow$, $\leftrightarrow$
quantifiers	$\forall$, $\exists$
alethic modal operators	$\Box$, $\Diamond$
tense operators	$\mathcal{F}, \mathcal{P}, \mathcal{G}, \mathcal{H}$
context-sensitive operators	$\mathcal{A}, \mathcal{N}$

Wffs and Sentences

1. If Π is an $\langle m, n\rangle$-place predicate and $i_1, ..., i_m$ are i-terms, and $p_1, ..., p_n$ are p-terms, then $\Pi i_1 ... i_m p_1 ... p_n$ is a wff.
2. if ϕ and ψ are wffs, then
   (a) $\neg\phi$ is a wff
   (b) $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \rightarrow \psi)$, $(\phi \leftrightarrow \psi)$ are wffs
   (c) $\forall\alpha\phi$ and $\exists\alpha\phi$ are wffs (α a p-variable or an i-variable)
   (d) $\Box\phi$, $\Diamond\phi$, $\mathcal{F}\phi$, $\mathcal{P}\phi$, $\mathcal{A}\phi$, $\mathcal{N}\phi$ are wffs.
3. Nothing else is a wff.

A sentence is a wff with no free variables.

4.2 IL Models

Perhaps the biggest difference between the IL models defined below and the ones Kaplan uses is that each model below has only a single context. This allows us to drop the mention of context from the definition of truth-in-a-model: with only one context per model, specifying the model is sufficient to specify the context.

A model is an ordered 6-tuple

$$\langle W, T, D, P, C, I\rangle$$

where:

1. W is a non-empty set of points (the set of worlds)
2. T is the set of integers (the set of times)[12]
3. D is a non-empty set (the domain of individuals)
4. P is a non-empty set (the domain of places)
5. C is a quadruple, $\langle a_C, p_C, n, @\rangle$ (the model's *context*[13] where:
   (a) $a_C \in D$ (the agent of the context)
   (b) $p_C \in P$ (the place of the context)
   (c) $n \in T$ (the time of the context, the "now")
   (d) $@ \in W$ (the world of the context, the "actual world")
6. I is a function from expressions to appropriate values for those expressions as follows:
   (a) If α is a non-logical i-constant, then $I(\alpha) \in U$
   (b) If α is a non-logical p-constant, then $I(\alpha) \in P$
   (c) If Π is a $\langle m, n\rangle$-place non-logical predicate, then $I(\Pi)$ is a function such that for each $t \in T$ and $w \in W$, $I(\Pi)(t, w) \subseteq (D^m \times P^n)$ (i.e. a function from time-world pairs to an sequence consisting of an m-tuple of individuals from D followed by an n-tuple of places from P—the predicate's *intension*.)
7. (a) $o \in D$ iff $\exists t \in T$ and $\exists w \in W$ such that $o \in I_{Ex}(t, w)$
   (b) $\langle a_C, p_C\rangle \in I_{Loc}(n, @)$
   (c) For all $j \in D$ and $k \in P$, if $(j, k) \in I_{Loc}(t, w)$, then $j \in I_{Ex}(t, w)$

4.3 Variable Assignments

Variable assignments need to accommodate the fact that we have two kinds of variable, which take their denotations from D and P respectively. So let f_i be a function taking each i-variable to a value in D, and f_p a function taking each p-variable to a value in P. Then $f_i \cup f_p$ is variable assignment. If g is a variable assignment then g^{ξ}_{o} is the variable assignment just like g except that its value for the variable ξ is o.

4.4 Truth and Consequence

Truth and Denotation (in M at (t, w) on g)

We need denotations for several kinds of term: i- and p-variables, i-and p-constants, and our logical terms $\mathfrak{i}$ and $\mathfrak{h}$. Both the denotation of a term and the truth-value of a sentence can depend on a lot of things: the model, an

[12]Note that the set of integers has its own ordering relation and we will exploit this in the truth-conditions for the tense operators (rather than adding a separate ordering relation to the model.) This is a trick for keeping the models simpler than they might otherwise be.

[13]We call the time and the agent of the context n and @ to emphasise continuity with tense and modal models which contain a single privileged time and world used to define truth in the model—also thought of intuitively as "the now" and the "the actual world".

assignment, and the time and world. Of course, the truth-value of context-sensitive wffs depends on the context as well, but since IL-models (unlike LD-models) have a single context per model, fixing the model itself is sufficient to fix the context.

We write $[\alpha]_{Mtwg}$ for the **denotation** of the term α in the model M, at the time t in the world w on the assignment g, and $V_{M_g}(\phi, t, w)$ for the **truth-value** of the sentence ϕ at a time t and world w, in the model M on an assignment g.

1. $[\alpha]_{Mtwg} = \begin{cases} I(\alpha), & \text{if } \alpha \text{ is a non-logical } i\text{-constant or } p\text{-constant} \\ g(\alpha), & \text{if } \alpha \text{ is an } i\text{-variable or } p\text{-variable} \\ a_C, & \text{if } \alpha \text{ is } \mathfrak{i} \text{ } (\mathfrak{i} \text{ refers to the model's agent}) \\ p_C, & \text{if } \alpha \text{ is } \mathfrak{h} \text{ } (\mathfrak{h} \text{ refers to the model's place}) \end{cases}$

2. Where Π is an $\langle m, n\rangle$-place non-logical predicate, $i_1, ...i_m$ are i-terms and $p_1, ...p_n$ are p-terms, then

   $V_{Mg}(\Pi i_1..i_m p_1..p_n, w, t) = 1$ iff $\langle [i_1]_{Mtwg}, \ldots, [i_m]_{Mtwg}, [p_1]_{Mtwg}, \ldots, [p_n]_{Mtwg}\rangle \in I_\Pi(t, w)$

3. (a) Where α and β are i-terms

   $$V_{M_g}(\alpha =^{\langle 2,0\rangle} \beta, t, w) = 1 \quad \text{iff} \quad [\alpha]_{Mgtw} = [\beta]_{Mgtw}$$

   (b) Where α and β are p-terms

   $$V_{M_g}(\alpha =^{\langle 0,2\rangle} \beta, t, w) = 1 \quad \text{iff} \quad [\alpha]_{Mgtw} = [\beta]_{Mgtw}$$

4. (a) $V_{M_g}(\neg\phi, t, w) = 1$ iff $V_{M_g}(\phi, t, w) = 0$
   (b) $V_{M_g}(\phi \wedge \psi, t, w) = 1$ iff $V_{M_g}(\phi, t, w) = 1$ and $V_{M_g}(\psi, t, w) = 1$

   $\vdots$ etc.

5. (a) If ξ is an i-variable, then
   $V_{M_g}(\forall\xi\phi, t, w) = 1$ iff for all elements $o \in D, V_{M_{g^\xi_o}}(\phi, t, w) = 1$
   (b) If ξ is a p-variable, then
   $V_{M_g}(\forall\xi\phi, t, w) = 1$ iff for all elements $p \in P, V_{M_{g^\xi_p}}(\phi, t, w) = 1$

6. (a) $V_{M_g}(\mathcal{F}\phi, t, w) = 1$ iff there is $u \in T, t < u$ and $V_{M_g}(\phi, u, w) = 1$
   (b) $V_{M_g}(\mathcal{G}\phi, t, w) = 1$ iff for all $u \in T$ such that $t < u$, $V_{M_g}(\phi, u, w) = 1$
   (c) $V_{M_g}(\mathcal{P}\phi, t, w) = 1$ iff there is $u \in T, u < t$ and $V_{M_g}(\phi, u, w) = 1$
   (d) $V_{M_g}(\mathcal{H}\phi, t, w) = 1$ iff for all $u \in T$ such that $u < t$, $V_{M_g}(\phi, u, w) = 1$

7. (a) $V_{M_g}(\Diamond\phi, t, w) = 1$ iff there is $v \in W$ such that $V_{M_g}(\phi, t, v) = 1$
   (b) $V_{M_g}(\Box\phi, t, w) = 1$ iff for all $v \in W$, $V_{M_g}(\phi, t, v) = 1$

8. (a) $V_{M_g}(\mathcal{A}\phi, t, w) = 1$ iff $V_{M_g}(\phi, t, @) = 1$
   (b) $V_{M_g}(\mathcal{N}\phi, t, w) = 1$ iff $V_{M_g}(\phi, n, w) = 1$

Truth at a time and a world in a model:
A sentence ϕ is true in a model M at a time t and a world w, if $V_{M_g}(\phi, t, w) = 1$

for all variable assignments g.

Truth in a model:
ϕ is true in a model M iff $V_M(\phi, n, @) = 1$. We write: $V_M(\phi) = 1$.

Note that we use the time, n, and world, $@$, *of the context* (C) for defining truth in a model.

A *set* of sentences Γ is true in a model M iff $V_M(\gamma) = 1$, for all $\gamma \in \Gamma$. We write: $V_M(\Gamma) = 1$.

Logical Consequence:
$\Gamma \vDash_{IL} \phi$ iff for all IL-models M, if $V_M(\Gamma) = 1$, then $V_M(\phi) = 1$.

5 An Application

The plan for the rest of this paper is to use IL to do some philosophy. Since it wouldn't be possible to do this work without a logic like IL, this will function as an argument for the usefulness of the logic. Our general topic is *barriers to entailment*, where a barrier to entailment is a thesis that says that it is impossible to get conclusions of a certain kind from premises of another. Perhaps the most famous barrier is Hume's Law, which says that you can't get an *ought* from an *is* or more carefully, that normative conclusions never follow logically from premise sets that are purely descriptive. Hume's Law is controversial, but other barriers are regarded as philosophical platitudes: you can't get universal conclusions from merely particular premises, or conclusions about the future if all your premises are about the past, or conclusions that say how things *must* be from premises that merely tell you how they actually are. The question I want to consider now is: is there an indexical barrier to entailment, i.e. one that says that it is impossible to get indexical sentences from premises that don't contain indexicals?

I'll start by talking a bit about why anyone might think there would be, then raise some problems and potential counterexamples. Then I'll switch gears to look at how we might use tense logic to prove the past/future barrier, and finally adapt that technique so that we can use our indexical logic to prove a (somewhat restricted) version of an indexical barrier.[14]

5.1 Motivation

There is informal work by several philosophers—especially Bar-Hillel, Casteñeda, Perry, and David Lewis—that is suggestive of an indexical barrier.[15] Here is Lewis' *Two Gods* case:

[14]This section is a summary of work presented more extensively in Russell (2022) and interested readers can learn about it in more detail there.

[15]Bar-Hillel (1954); Castañeda (1968); Perry (1979); Lewis (1979). See also Strawson: "The same sentence in different mouths may be used to make one true, and one false, statement

> There is a world with two gods, one nice, one nasty. The nice god lives on the tallest mountain and showers mana on the people. The nasty god lives on the coldest mountain and hurls thunderbolts. Being gods, they don't come to know things through normal animal perception—they don't see, or hear, or feel—but they are omniscient anyway in a distinctively godly way: each knows the truth-values of every non–indexical sentence. For example, they know that *There are two gods* and *The nice god lives on the tallest mountain* are true, and that *The nasty god lives on the tallest mountain* is false. They don't, initially, know the truth-values of indexical sentences, such as *I am the nice god* or *it is raining here.* Our question is: can either god deduce indexical sentences from the constant sentences they already know to be true?[16]

Lewis' comments suggest perhaps not.[17] One reason to think that might be right is that each of the two gods can draw on the same set of premises, though the truth-value of indexical sentences varies depending on which god you are; *I am the evil god* is true for the evil god, false for the good one. False sentences are not entailed by true premises, so indexical sentences are not entailed by the set of all true constant sentences. So far, so promising. Perhaps, now that we have an indexical logic, we can go on to *prove* an indexical barrier theorem as a metatheorem about that logic. But the indexical logic also clarifies some problems for the indexical barrier thesis.

5.2 Putative Counterexamples

The logic of indexicals allows us to formulate some putative counterexamples to an indexical barrier.[18]

0a. $p \vDash \mathcal{A}p$
b. $p \vDash \mathcal{N}p$
c. $p \vDash \mathcal{N}Loc(\mathfrak{i}, \mathfrak{h})$
d. $\forall x Bx \vDash B\mathfrak{i}$

The first two of these are famous features of Kaplan's logic, shared with IL, and things are even slightly worse than they seem, since in each case the indexical

("My cat is dead") [...] It is also an unavoidable feature of any language we might construct to serve the same general purposes." Here Strawson seems to suggest that indexical language is in some sense ineliminable—or "essential". (Strawson, 1952, 212)

[16]Slightly adapted from (Lewis, 1979, 520-521)

[17]I hedge because his set up is slightly different from mine. But on 521 he writes: "[n]either one knows whether **he** lives on the tallest mountain or on the coldest mountain, nor whether **he** throws manna or thunderbolts." (bold added.)

[18]I take some slight liberties with presentation of these counterexamples for the sake of readability. Officially there is no sentence letter p or q in the formal language presented above, and so the official version should have an atomic sentence like Ba or Cb in their place. But this presentation increases the readability of the counterexamples without creating any deep problems.

conclusion is logically equivalent to the premise, not merely entailed by it. 0c. also exploits a famous feature of Kaplan's logic: the formalisation of *I am here now* as the logical truth. Since logical truths are entailed by all sets of premises, they are entailed by all sets of constant sentences. 0d. involves quantifiers, and there are some other variants on this. We might also consider e.g. $\neg\exists x Bx \vDash \neg B\mathfrak{i}$.

Does this close the question of whether there is an indexical barrier? What would become then of the intuitive idea that the values of indexical sentences vary with context, and this explains why they aren't entailed by sentences which don't? I am going to suggest that in fact the counterexamples above can help us to refine and improve the formulation of an indexical barrier into a true precisification of the intuitive idea. In order to make that case, I will start by considering a related barrier to entailment: the past/future barrier.

5.3 The Past/Future Barrier

A.N. Prior, the founder of tense logic, is also well-known for his objection to a further barrier to entailment—Hume's Law—against which he argued that if the disjunction $p \vee \mathcal{O}q$ is normative, then $p \vDash p \vee \mathcal{O}q$ is a counterexample, and if the disjunction is not normative, then $p \vee \mathcal{O}q, \neg p \vDash \mathcal{O}q$ will serve instead.[19] It is much less well-known that Prior also argued that the development of tense logic allowed us to see that the past/future barrier was false as well. He cites Bennett, who calls the barrier "a splendid discovery" and went on "philosophers have known about it since Hume and, if they had paid attention, they could have known about it since Leibniz." (Bennett, 1961, 61) Prior responds:

> One thing that the development of tense-logic makes quite clear—if it was not clear before—is that this alleged 'discovery' is in fact a falsehood. (Prior, 1967, 57))

Here is a version of Prior's counterexample transposed into the language of IL:

1a. $\mathcal{P}p \vDash \mathcal{F}\mathcal{P}p$

Rendered informally, this says that "At some time in the past p" entails "at some time in the future, at some time in the past p." The premise appears to be about the past and the conclusion about the future (it makes a claim about the future, and contains the $\mathcal{F}$-operator.) The argument is valid in IL (as in Prior's system) and so this appears to be a counterexample to the claim that no set of premises only about the past entails a claim about the future.

The situation is perhaps even worse than Prior makes it seem. Here are some other putative counterexamples to the past/future barrier in the formalism of tense logic:

1b. $p \vDash p \vee \mathcal{F}q$
c. $p \vDash \mathcal{G}(q \vee \neg q)$

[19] Prior (1960) presents this argument informally.

d. $p, \neg p \vDash \mathcal{F}q$
e. $a = b \vDash \mathcal{F}(a = b)$

We might begin to formulate a response on behalf of Hume's (and perhaps Leibniz') barrier by observing that Hume's point was really restricted to future *matters of fact*; he would have allowed statements expressing relations of ideas. In a similar spirit, our formulation of the past/future barrier can allow logical truths—as in 1c.—to follow from premises about the past. The real past/future barrier concerns what we might call *logically synthetic* future sentences—those true in some models, false in others.

5.3.1 Future-switching

Why think there is a past/future barrier at all? Here is a reason that resembles the earlier motivation for the indexical barrier: future synthetics (that is, synthetic sentences concerning the future) are sensitive to *changes to the future*: on some models of the future they are true, on others false. Sentences about the past, by contrast, are not sensitive in this way; if the past stays the same, but the future changes, then future synthetics could go false, while the past sentences retain whatever truth-values they had. But that suggests that the past sentences could be true while the future ones are false, in which case the past sentences wouldn't entail the future sentences.

We can use this (somewhat hazy) thought as the inspiration for a precisification of what we mean by future sentences. The first step is to get clearer about what is meant by *changing the future*. The intuitive idea will be that future-switching (as we will call it) involves changing the I-function values of the primitive non-logical expressions for world-time pairs consisting of the actual world paired with a time later than the present moment (n.) The way we will implement this in our IL models involves *swapping* the I-values from some possible future for the values in the actual future. The complexity of IL models makes it hard to draw diagrams representing all their different aspects. But we can represent a part of the model—the values of the I-function for world-time pairs—in a table like the one below.

	w	@	**u**
$\mathbf{t_5}$	p	p, q, r	
$\mathbf{t_4}$	q	p, q	p
$\mathbf{n}$	p	p	
$\mathbf{t_2}$	p		
$\mathbf{t_1}$	p	p, q	

Figure 1: Model M: $\mathcal{F}(q \wedge r)$ is true in this model.

The rows of the table represent times (time's arrow flies upwards) and the columns represent worlds. Row n is the present moment, column @ the actual world of the model (both members of the context C.) Each cell of the table represents a world-time pair, and in it we write (all and only) the sentence letters to which I assigns the value 1, relative to the world-time pair represented by that cell.[20] Since truth-in-a-model is truth at the model's "actual now", or $\langle @, n \rangle$, the diagram above represents a model, M, in which the sentences p, $\mathcal{P}p$, $\mathcal{F}r$, $\mathcal{F}(q \wedge r)$, and $\mathcal{F}\mathcal{P}q$ are all true.

Each partial column above the n-row of the table represents a (sub) model of the future, with the area shaded grey representing the model's *actual future*. An easy visual way to think about the intermodel relation of future-switching is as swapping the model's actual future for one of the model's "possible futures." i.e. we might swap the partial column above n in column u with the actual future, to get a new model that stands in the future-switch relation to the old:[21]

	w	@	**u**
$\mathbf{t_5}$	p		p, q, r
$\mathbf{t_4}$	q	p	p, q
$\mathbf{n}$	p	p	
$\mathbf{t_2}$	p		
$\mathbf{t_1}$	p	p, q	

Figure 2: Model N: $\mathcal{F}(q \wedge r)$ is false in this model.

That's the intuitive idea. We can now define this more carefully:

Definition (Future-switching)**.** *For all models M and N, N is a future-switch of M, where $M = \langle W, T, D, P, C, I \rangle$, iff*

1. $N = \langle W, T, D, P, C, I^* \rangle$
2. *for all primitive non-logical expressions p, there is exactly one $w \in W$ such that for all $t \in T$*

[20] Even these tables simplify slightly in that they represent the values of sentence letters, whereas IL computes the values of atomic sentences from those of terms and predicates. So we should really represent the extensions of predicates for each world and time. (If you like, think of each cell of the table as containing a little first-order model of its own, giving the extensions of each predicate—though each cell agrees on the extensions of individual non-logical constants—those are rigid across times and worlds.) Here I think the simplification is worth it for the accessibility of the initial exposition.

[21] One might reasonably wonder why one is not simply allowed to edit the actual future however one likes (rather than swapping it for a different future from the same model.) The reason is to do with retaining the intuitive status of $\Box p$ as non-Future. See Russell (2022) for more details.

(a) if $n < t$,

$$I^*(p)(t, w) = I(p)(t, @)$$
$$I^*(p)(t, @) = I(p)(t, w)$$[22]

(b) for $t \leq n$,

$$I^*(p)(t, w) = I(p)(t, w)$$

3. and for all $u \in W$ where $u \neq w$ and $u \neq @$, and all $t \in T$,

$$I^*(p)(t, w) = I(p)(t, w)$$

We use this future-switching relation on models to define the class of Future sentences:

Definition (Future sentence)**.** *A sentence ϕ is* Future *if it is future-switch* breakable*, i.e., there are models M, N such that N is a future-switch of M and ϕ is true in M but not in N.*[23]

Remark. *Future sentences then include: $\mathcal{F}p$, $\mathcal{G}p$, $\mathcal{FP}p$, $\mathcal{P}p \vee \mathcal{F}q$, $\mathcal{P}p \rightarrow \mathcal{F}q$.*

Remark. *Non-Future sentences include: $p, \mathcal{P}p, \mathcal{H}p$, as well as unsatisfiable sentences—like $\mathcal{F}p \wedge \neg\mathcal{F}p$—and logical truths like $\mathcal{F}p \rightarrow \mathcal{F}p$ or $\mathcal{P}p \rightarrow \mathcal{GP}p$.*

These definitions take care of the putative counterexamples 1c) and 1e). In both the conclusion is not Future, since it is not future-switch breakable. But we are still left with three valid arguments from non-Future sets of sentences to Future conclusions:

1a. $Pp \vDash FPp$
b. $p \vDash p \vee \mathcal{F}q$
d. $p, \neg p \vDash \mathcal{F}q$

It's worth noting, at this point, that FPp and $p \vee Fq$ share an interesting feature: each can be made true in two different ways. For example, $p \vee Fq$ can be true in a model because p is true, or because (just) Fq is true. FPp can be true in a model because p is true in the past (or present) or instead because p is true in the future (and not in the past or the present.) As we might put it: each can be made true by a fact about the future, or alternatively by a fact about the past or the present. Of course, if it is made true by a fact about the past or the present, future-switching will not make it false, because it won't change that past or present fact. These sentences are only future-switch breakable because they can also be made true by a fact about the future alone, and future-switching can destroy those facts.

[22] Here $I(p)$ is the intension assigned to the expression p by the I-function, and $I(p)(t, w)$ is the extension of that expression (given the intension $I(p)$) at the time-world pair t, w.

[23] We retain the initial capital in *Future* as a reminder that this is a term of art. For reasons I don't have space to go into here we'll extend this definition to sets of sentences, i.e. a set of sentences is Future if there is a model of the set that makes it true, and a future-switch of the model that makes it false (i.e. makes at least one sentence in the set false.) Note that it is possible for every sentence in a set to be Future without the set being Future, e.g. $\{\mathcal{F}p, \neg\mathcal{F}p\}$.

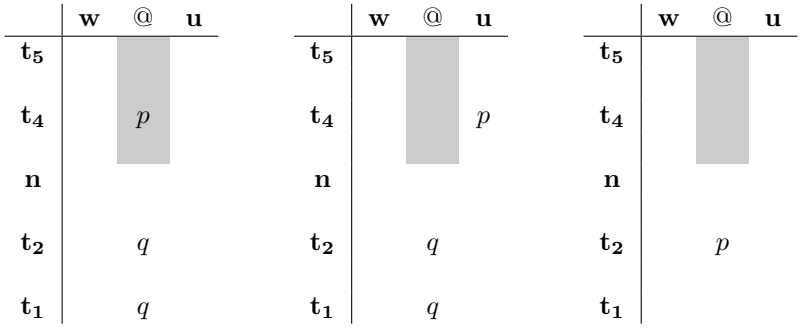

Figure 3: A model (left) that makes FPp true and a future-switch of that model (center) that makes it false. In addition a model (right) which makes FPp true but which has no future-switches that make it false.

In each of 1a. and 1b. the truth of the premise ensures that we are in one of the models which leaves the conclusion invulnerable to future-switching. For example, making Pp true ensures that p is true in the actual past, and so ensures that future-switching won't change the truth-value of the conclusion FPp. Similarly, making p true ensures that we are in a model which has no future-switches that make $p \vee Fq$ false. Another way to think of this: in each case, though the conclusion ϕ is, quite generally, susceptible to having its truth-value changed by future-switching, the set consisting of the union of the premises, Γ, with the singleton of the conclusion, $\{\phi\}$ is not susceptible to future-switching: you can only make the entire set true in a model which has no future-switches that make it false. We can exploit this feature to formulate a *limited* barrier to entailment as a metatheorem about IL:

Theorem (Past/Future Barrier)**.** *Let Γ be a set of premises which is not Future, and ϕ a sentence which is Future. Then $\Gamma \nvDash_{IL} \phi$, unless $\Gamma \cup \{\phi\}$ is not Future.*

It is straightforward to prove this by reasoning about IL models.

Proof. *Suppose ϕ is Future, Γ is not Future, and moreover $\Gamma \cup \{\phi\}$ is Future. We show $\Gamma \nvDash_{IL} \phi$. Since $\Gamma \cup \{\phi\}$ is Future it is future-switch breakable, i.e. satisfiable by some model M which has a future-switch N that does not satisfy it. But Γ is not future-switch breakable, and so is true in N. But N (since it does not satisfy the union of Γ and $\{\phi\}$) does not satisfy ϕ, and hence N is a countermodel showing $\Gamma \nvDash_{IL} \phi$.* □

Here is how the theorem avoids each of the putative counterexamples. 1a. is not a counterexample because the union of the premises and conclusion, $\{\mathcal{P}p, \mathcal{F}\mathcal{P}p\}$, is not Future. 1b. is not a counterexample because $\{p, p \vee \mathcal{F}q\}$ is not Future. 1c. is not a counterexample because the conclusion, $\mathcal{G}(q \vee \neg q)$, is not Future. 1d. is not a counterexample because $\{p \wedge \neg p, \mathcal{F}q\}$ is unsatisfiable and so not Future. 1e. is not a counterexample, because $\mathcal{F}(a = b)$ is not Future-switch breakable and hence not Future on our definitions.

5.4 Back to Indexicals

We proved a version of the past/future barrier above by getting clearer about what was meant by *Future*. Future sentences were those that could be made false by future-switching—they were future-switch breakable. In order to apply the same ideas to the other barriers, we need to think of the sentences that appear in the conclusions as breakable with some kind of change to an appropriate model. For example, it is a familiar idea to many linguists that Universal sentences are those that can break when we *extend the domain* of the model. (Add something "not-B" to a model in which everything is "B" and $\forall x Bx$ goes from true to false. It is *domain-extension breakable*.) Normative sentences might be breakable with changes in the normative standards, necessity-attributing sentences breakable with the addition of new possible worlds, and indexical sentences breakable with context-shifts.

We can generalise the limited Past/Future barrier above to a Limited General Barrier:[24]

Theorem (Limited General Barrier Theorem)**.** *Let R be a binary relation on a set of models U. If ϕ is R-breakable over U but Γ is not, then $\Gamma \not\models \phi$ unless $\Gamma \cup \{\phi\}$ is not R-breakable (over U).*

Proof. *Suppose ϕ is R-breakable, Γ is not R-breakable, and moreover $\Gamma \cup \{\phi\}$ is R-breakable. We show $\Gamma \not\models \phi$. Since $\Gamma \cup \{\phi\}$ is R-breakable it is satisfiable by some model M and there is a model N, MRN, and N does not satisfy $\Gamma \cup \{\phi\}$. But Γ is not R-breakable, so there is no model R-related to M in which Γ is not true. So N (since it does not satisfy the union of Γ and $\{\phi\}$) does not satisfy ϕ, and hence N thus a countermodel to $\Gamma \models \phi$.* □

We could then get an Indexical barrier by specifying an appropriate R relation: context-shifting. We will do this in a moment, but there is one respect in which we need to be careful about context-shifts. The intuitive idea will be that a sentence counts as Indexical just in case changing the context (and nothing else) can sometimes change its truth-value from true to false.[25] The problem arises from the fact that contexts contain worlds and times, which are also aspects of the circumstances of evaluation. Suppose we change the context, $\langle a_C, p_C, n, @\rangle$ by altering the value of @. Which sentences could change their truth-values as a result of such a change to the context? Answer: *any contingent sentence.* This will include sentences like *snow is white*, and not merely sentences containing overt modal indexicals. Similarly, if we change the time, this will threaten any "temporally contingent" sentence. Intuitively, we want to alter the context so that indexical sentences come to express propositions that have different truth-values. But when we alter the time and world in the Kaplanian framework, we can also alter the values of some propositions. The solution

[24] Robbie Williams suggested that since I live in Australia now I should call this the *Great Barrier Theorem*, but I fear the appearance of cockiness might outlive the joke.

[25] Context-shifting is—like future-switching—symmetric, so we can abbreviate this to "can sometimes change its truth-value." The same is not true for domain-extension—the relation used to capture Universality—however.

is to limit ourselves to a *partial* context shift: we can change the elements of the context of use *that are not also elements of the circumstance of evaluation.* In the present framework, that amounts to changing the agent or the place of the context—but not the world or the time.

Definition (Partial Context-Shifting)**.** *An IL model N is a partial context-shift of a model M just in N is exactly like M except (possibly) for the values of a_C and p_C.*

We then define an Indexical sentence as one breakable by (partial) context-shifting.

Definition (Indexical Sentence)**.** *A sentence ϕ is Indexical just in case it is partial context-shift breakable, that is there is model M and a model N such that ϕ is true in M, ϕ is false in N, and N is a partial context-shift of M.*

Examples of Indexical sentences: $B\mathfrak{i}$, $C\mathfrak{h}$, $B\mathfrak{i} \vee Da$, $B\mathfrak{i} \rightarrow Da$.

Examples of Non-Indexical sentences: Ba, $\mathcal{N}Ba$, $\mathcal{A}Ba$, $B\mathfrak{i} \wedge \neg B\mathfrak{i}$, $Ba \vee \neg Ba$, $Loc(\mathfrak{i}, \mathfrak{h})$, $\mathcal{N}Loc(\mathfrak{i}, \mathfrak{h})$.

The last one is an important case. *I am here now* might contain indexical expressions, but the whole sentence is not Indexical on this criterion: changing the context never changes its truth-value. So like other logical truths (and like $Pp \rightarrow FPp$ for Future-sentences) it is classified as non-Indexical on the present taxonomy. We can now formulate our Limited Indexical Barrier theorem.

Theorem (Limited Indexical Barrier Theorem)**.** *If ϕ is Indexical but Γ is not, then $\Gamma \not\models \phi$ unless $\Gamma \cup \{\phi\}$ is not Indexical.*

Proof. *Letting R be the relation of partial context-shifting, this is an instance of the Limited General Barrier Theorem.* □

Here is how the theorem handles the putative counterexamples. 0a. and 0b. are not counterexamples because $\mathcal{A}p$ and $\mathcal{N}p$ are not partial context-shift breakable, and so not Indexical in our sense. 0c. is not a counterexample because $\mathcal{N}Loc(\mathfrak{i}, \mathfrak{h})$, being a logical truth, is not partial context-shift breakable, and so not Indexical in this sense. And 0d. is not a counterexample, though the premise is non-Indexical and the conclusion Indexical, because the set $\{\forall x Bx, B\mathfrak{i}\}$ is not Indexical—it meets the theorem's "unless" clause.

6 On the significance of the formal system

So we have (at least) these two things from "Demonstratives": a picture of how indexicals work and a logic, LD. The former is highly influential and supports revised conceptions of the logical properties, the latter was called by Kaplan "the most important part and certainly the most convincing part" (487) of his work. But they are different kinds of thing, and the situation reminds

me of Kripke's comment in *Naming and Necessity* that he "wasn't going to present an alternative theory" of names because "philosophical theories are in danger of being false" and instead he'll present "just a *better picture*". (Kripke, 1980, 64, 93) These terms—*picture* and *theory*—weren't being used in any very technical senses, but it's clear that theories are more likely to be false, and so it seems likely that they are more definite, more detailed, more informative. Pictures are simpler, "higher level", and we might be tempted to reach for the word *conceptual* to describe this kind of work. Kaplan calls his own picture *metaphysical*, and we could call it *abstract*, but then (we might also find ourselves thinking) even LD is a *logic* and what is more abstract than that?

One thing that I think we can say, is that LD is one attempt to make the key ideas from the metaphysical picture more definite and precise, and to fit them into a bigger story (logical model theory, and perhaps formal semantics) in a way that shows us how to connect the picture to a great deal of other work. One way in which LD is more definite is that it makes decisions about issues that the picture leaves unsettled. Can propositions change their truth-values over time? Is *here* analogous to *now* or to *I*? Is *actually* really an indexical? What kinds of things can serve as referents of *I*? (Do they need to be persons or will any element of the domain do?) Can *I* fail to have a referent, like an empty name? Are there contexts with respect to which *I am here* is false? These issues are extraneous to the picture of indexicals Kaplan laid out: we could keep the picture and go either way on the answers to the questions, because the picture can be instantiated in lots of different ways, as part of many different frameworks. And this is one of the reasons that LD almost inevitably gets some things wrong: for LD to be right it would have to be right on *so many things.*

But the reason LD matters is absolutely not that it gets every detail right. The logic that I used in the previous section is a *variant* on LD—not LD itself—but I don't see this as any indictment of LD, except to the extent that it wasn't quite the right tool for the task at hand—a pragmatic issue, not a theoretical one. What the two have in common is that they realise Kaplan's picture for indexicals in a precise, definite way—in this case a language with a set of models. Both distinguish character from content, and sensitivity to context from sensitivity to circumstance of evaluation, e.g. indexicality from non-rigidity.[26] Both treat indexicals as logical expressions and allow for the establishment of contingent logical truths—$A p \leftrightarrow p$ and $\mathcal{N} Loc(\mathfrak{i}, \mathfrak{h})$—sentences whose truth-value is guaranteed by the meanings of those indexical expressions, even if the proposition expressed in any one context might turn out to be a contingent one. Both allow the validity of arguments containing indexicals to be established by reasoning about models.

I think a key point here is that the process of constructing more definite, more precise, instantiations like this has a tendency to expose tensions, problems, and unclarities in the original picture. That is part of why it is a useful method, and why the first success in constructing a logic for indexicals mattered.

[26] For clarity I should at least mention that neither logic captures *every* feature of the picture—for example, they aren't fine-grained enough to distinguish rigidity from direct reference. But both instantiate the picture using model theory.

> . . . we can often produce mathematical models of fragments of philosophy and, when we can, we should. No doubt the models usually involve wild idealisations. It is still progress if we can agree what consequences an idea has in one very simple case. Many ideas in philosophy do not withstand even that elementary scrutiny because the attempt to construct a non-trivial model reveals a hidden structural incoherence in the idea itself. (Williamson, 2008, 291)

Until Kaplan's formal system, it remained an open question: is this thing actually going to *work*? LD was a vindication of the picture it instantiates—a demonstration that a language *could* function as the picture describes, and (even better) in a way that fits with other entrenched frameworks. The construction of LD functioned as a kind of proof of possibility for the underlying picture, and it showed how to unify that picture with existing frameworks.

LD's definiteness is also useful for the same reason definiteness is useful elsewhere in other parts of logic: "definite, precisely formulated formal systems" are necessary conditions for the possibility of metatheory. We can't *prove* things about vague, informal pictures. I hope my own work goes some way towards showing that metatheory for indexical logics can be used to increase understanding of philosophical issues; for example, it can help us see what underlies the various intuitions about the so-called "essentiality" of the indexical brought out in the work of philosophers like Bar-Hillel, Casteñeda, Perry, and Lewis, and how this relates to other barriers to entailment. But this too can function as a kind of proof of possibility: this time of the fruitfulness of the formal system from "Demonstratives".

References

Bar-Hillel, Y. (1954). Indexical expressions. *Mind*, 63(251):359–379.

Bennett, J. (1961). A myth about logical necessity. *Analysis*, 21(3):pp. 59–63.

Berg, J. (1991). Themes from Kaplan (review). *International Studies in Philosophy*, 23(3):92–94.

Braun, D. (2017). Indexicals. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2017 edition.

Castañeda, H.-N. (1968). On the logic of attributions of self-knowledge to others. *Journal of Philosophy*, 65(15):439–456.

Cottingham, J., Stoothoff, R., and Murdoch, D., editors (1984). *The Philosophical Writings of Descartes*. Cambridge University Press, Cambridge.

Forbes, G. (1989). Indexicals. In *Handbook of Philosophical Logic*, volume IV, chapter 6. D Reidel Publishing Company, Dordrecht.

Georgi, G. (2022). Indexicals and demonstratives. *Internet Encyclopedia of Philosophy*. (n.b. The IEP lists only the date viewed.).

Goldfarb (2003). *Deductive Logic*. Hackett, Indianapolis, IN.

Jeffrey, R. (1967). *Formal Logic: Its Scope And Limits*. McGraw Hill Book Company, New York.

Kaplan, D. (1989a). Afterthoughts. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes from Kaplan*. Oxford University Press, New York.

Kaplan, D. (1989b). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes from Kaplan*. Oxford University Press, New York.

Kripke, S. A. (1980). *Naming and Necessity*. Blackwell, Oxford.

Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, 88:513–543.

Newman, L. (2019). Descartes' epistemology. *The Stanford Encyclopedia of Philosophy*. Spring Edition, Edward N. Zalta (ed).

Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13:3–21.

Predelli, S. (1998). I am not here now. *Analysis*, 58(2):107–115.

Predelli, S. (2008). "I" exist: The meaning of "I" and the logic of indexicals. *American Philosophical Quarterly*, 45(1):57–65.

Prior, A. N. (1960). The autonomy of ethics. *The Australasian Journal of Philosophy*, 38:199–206.

Prior, A. N. (1967). *Past, Present and Future*. Clarendon Press, Oxford.

Quine, W. V. O. (1950). *Methods of Logic*. Holt, Rinehart and Winston, New York.

Quine, W. V. O. (1986). *Philosophy of Logic*. Harvard University Press, Cambridge, Mass.

Radulescu, A. (2015). The logic of indexicals. *Synthese*, 192:1839–1860.

Russell, G. K. (2022). How to prove Hume's law. *Journal of Philosophical Logic*, 51:603–632.

Salmon, N. (2002). Demonstrating and necessity. *The Philosophical Review*, 111(4):497–537.

Strawson, P. F. (1952). *Introduction to Logical Theory*. Metheun and Company Ltd, London.

Strawson, P. F. (1985/1950). On referring. In Martinich, A. P., editor, *The Philosophy of Language*. Oxford University Press, Oxford, 4th edition.

Vision, G. (1985). I am here now. *Analysis*, 45(4):198–199.

Williamson, T. (2008). *The Philosophy of Philosophy*. Blackwell, Oxford.